Patterns of population genetic structure among Australian and South Pacific humpback whales (*Megaptera novaeangliae*)

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in Biological Sciences,

The Australian National University, Canberra, in collaboration with the Australian Marine Mammal Centre, Hobart

By

Natalie Tara Schmitt

B. Sc. (Hons) in Zoology and Botany

University of Western Australia, Perth

October 2012

©David Donnelly

There Leviathan,

Hugest of living creatures, on the deep

Stretch'd like a promontory sleeps or swims,

And seems like a moving land; and at his gills

Draws in, and at his breath spouts out a sea.

John Milton, *Paradise Lost,* quoted in title page to the first, English edition of *Moby-Dick*

DECLARATION

The research presented in this thesis is my own original work except where due reference is given in the text. All the chapters were the product of investigations carried out jointly with others but in all cases I am the principal contributor to the work. No part of this thesis has been submitted for any previous degree.

………………………………………………..

Natalie Tara Schmitt

October, 2012

THESIS PLAN

The dissertation is written as a series of four self-contained chapters ready for publication plus a concluding chapter. All chapters are written in the format of the target journal with tables, figures and appendices included at the end of each chapter. As such the pronoun “we” is used to represent co-authors of material intended for publication. Below I outline the contributions of my co-authors.

Chapter 1 - Low levels of genetic differentiation characterize Australian humpback whale (Megaptera novaeangliae) populations.

I was responsible for collecting the eastern Australian humpback skin samples, with Mike Double collecting the samples from western Australia. I was also responsible for conceptual development, laboratory work, statistical analysis and writing, with editorial and analytical assistance from Mike Double, Rod Peakall and Simon Jarman. External co-authors Scott Baker and Curt Jenner provided editorial assistance. I was also given advice in the lab by Simon Jarman and James Marthick.

Chapter 1 has been provisionally accepted for publication subject to revision in an international peer-reviewed journal (*Marine Mammal Science*). The chapter has also been presented at the International Whaling Commission Stock definition and Southern Hemisphere Scientific Sub-committee meetings held in Panama, 2012 (SC/64/SH15).

Schmitt, N.T., M.C. Double, C.S. Baker, D. Steel, K.C.S. Jenner, M.-N.M. Jenner, D. Paton, et al. In Press. Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations. Marine Mammal Science.

Chapter 2 - Mixed-stock analysis of humpback whales (Megaptera novaeangliae) on Antarctic feeding grounds.

The South Pacific Whale Research Consortium provided mtDNA and nuclear data from humpback whale breeding populations of the South Pacific. Mike Double, Simon Childerhouse, Rochelle Constantine, Nick Gales, Curt Jenner and Dave Paton assisted in the collection of skin samples from the Southern Ocean. Scott Baker also provided seven samples collected previously in the Southern Ocean. I was responsible for the conceptual development of this chapter as well as the laboratory work, statistical analyses and writing. Editorial advice on the manuscript was also given by Rod Peakall, Mike Double, Scott Baker and Simon Jarman.

Schmitt, N.T., M.C. Double, C.S. Baker, N. Gales, S. Childerhouse, A.M. Polanowski, D. Steel, G.R. Albertson, C. Olavarria, C. Garrigue, M. Poole, N. Hauser, R. Constantine, D. Paton, K.C.S. Jenner, S.N. Jarman, R. Peakall. Mixed-stock analysis of humpback whales (*Megaptera novaeangliae*) on Antarctic feeding grounds. This chapter is formatted as a manuscript for submission to *Marine Mammal Science*.

Chapter 3 – Re-assessing the genetic evidence for sex-specific migratory route choice in eastern Australian humpback whales (Megaptera novaeangliae)

I was responsible for most of the conceptual development and statistical analyses of this chapter with assistance from Rod Peakall and Mike Double. Genetic data was sourced from chapter 1 and chapter 2 with the Evan's Head samples collected, DNA extracted and sequenced by Andrea Polanowski and Simon Jarman. I received editorial assistance with the chapter from Rod Peakall and Mike Double.

Schmitt, N.T., M.C. Double, S.N. Jarman, A.M. Polanowski, R. Gales, R. Peakall. Re-assessing the genetic evidence for sex-specific migratory route choice in eastern

Australian humpback whales (*Megaptera novaeangliae*). This chapter is formatted as a manuscript for submission to *Marine Ecology Progress Series*.

Chapter 4 - Development and evaluation of single nucleotide polymorphism (SNP) markers for population structure analysis in the humpback whale (Megaptera novaeangliae).

In this chapter I was responsible for all intronic SNP discovery and Andrea Polanowski ascertained all exonic SNPs, with laboratory advice and assistance from Simon Jarman. I was also responsible for the conceptual design of the study, statistical analyses and writing, with editorial contributions from Simon Jarman, Mike Double and Rod Peakall.

Schmitt, N.T., A.M. Polanowski, M.C. Double, N. Gales, C.S. Baker, D. Steel, R. Peakall, et al. Small numbers of single nucleotide polymorphism (SNP) markers can detect population structure in the humpback whale (*Megaptera novaeangliae*). This chapter is formatted as a manuscript for submission to *Marine Mammal Science*.

Appendix – Development and application of SNPs for humpback whale population genetics (literature review).

This review was written at the onset of my PhD in 2008 as a way of informing myself and my supervisors of the available literature on SNP discovery and SNP genotyping in preparation for my fourth chapter. I was responsible for the collation of the literature and writing of the review. Simon Jarman, Mike Double and Rod Peakall provided some editorial contribution.

ACKNOWLEDGEMENTS

This project was a huge but rewarding challenge for me: despite having little experience in the field of molecular ecology I was inspired and motivated to be a part of the 'golden age' of genetic research, an age where we can now utilise genetic tools to learn more about elusive species like the humpback whale to inform management decisions. The Australian humpback whale also represents a suitable demographic and genetic model for the management of less tractable species of baleen whales and for the general study of gene flow among long-lived mobile vertebrates, due to their relatively high abundance and the ease with which they can be identified from natural markings. A symbol of global marine conservation, working with humpback whales has taught me what can be achieved in the field of population genetics with elusive species, and have greatly inspired me through the work of some amazing people I've had the privilege to meet and collaborate with over the course of the project.

This project would have been impossible to complete without the help of many people. Most of them are co-authors of my chapters or are named in the Acknowledgements section of each chapter. I would therefore like to use this section of the thesis to acknowledge those people that made great contributions to the journey through my PhD study and who are not necessarily colleagues.

I must first start by thanking my three supervisors. Their patience, understanding, encouragement and experience have been so valuable in my growth as a scientist. Thank you to Simon Jarman for your advice in the lab, particularly with the SNP discovery work; having come into this project with very little experience in genetics, your patience with me, expert guidance and ability to help resolve problems very quickly and with little hassle, as well as your friendly ear during times of frustration has made my experience in the lab a pleasurable one. My primary supervisor, Mike Double, thankyou

for firstly believing in my capabilities to undertake this project and thankyou especially for teaching me how to be a good and above all ethical scientist. Although your generous but tough approach was painful for me at times, I have learnt to demand good science from myself as well as from the scientific papers I read, after all, it is critical to get the science right when you're dealing with species conservation. And finally a special gratitude to my university supervisor, Professor Rod Peakall, your dedication to every one of your students and postdocs is astounding and I was never once left feeling like I was anything less than a top priority. You have always shown a great respect and belief in me and have instilled in me the skills and values that will continue to help me become a more rigourous scientist. You have helped make science exciting for me and taught me the joy of working in a positive, welcoming and professional atmosphere for research.

So keen was I to undertake this PhD, I volunteered in the genetics lab at the University of Tasmania for 6 months under the generous and patient guidance of Kevin Redd, who was doing his PhD at the time. My deep appreciate goes to him for teaching me the basics of genetic lab work and allowing me to make mistakes with his precious samples; I don't think I would have been permitted to take on the PhD without this experience.

My field work was definitely one of the most enjoyable aspects of the project. We are so privileged to be able to get up close to these magnificent animals. There were some essential people and organisations that were particularly supportive during this time. My appreciation to Jenny Robb, Scotty Sheehan, the Sapphire Coast Marine Discovery Centre and the Eden community for your support and interest in the project; it is so wonderful to have a community involved in conservation research and these guys really do value their whales. Thanks to Dave Paton for passing on his years of wisdom in the field and his ability to get me close enough to the whales that I simply could not miss

with a biopsy rifle, it was a joy to work with you. And to all the people involved in the Antarctic Whale Expedition in 2010, this was an absolute highlight thanks to an incredible team. I was so privileged to work with such an experienced group of people, where everyone put in an enormous effort to help me obtain enough samples for my study. I learnt a great deal about field work in extreme conditions on this voyage and I still pinch myself that I actually got the opportunity to visit this magical place and see humpback whales on their Southern Ocean feeding grounds, an experience I will never forget.

Throughout the journey, which has been incredibly tough at times, I was grateful to have the emotional support of some wonderful people. A big thankyou to all the other PhD students at the Australian Antarctic Division, SCUM as we called ourselves; there was both tears and laughter during our time together but your support will always be remembered and valued. My deep appreciation to my good friend Rebecca Leaper, your unwavering support, advice and friendship during this time was invaluable. I will never forget our many conversations on science and conservation, they were always tremendously inspiring and motivating. My family was also instrumental in their support and encouragement of me, particularly during the tough times. In particular, I want to thank my Grandmother for letting me live with her for two years and for being there for me when I needed a shoulder to cry on.

Finally, I want to thank my partner David Donnelly whom I met when he volunteered on my very first field trip off Eden. You have been so patient, supportive, unwavering and understanding and words simply cannot express my gratitude to you. Thankyou for being my rock.

And of course, I cannot forget my study species….it was never difficult to be inspired by you.

ABSTRACT

Humpback whales undertake long-distance seasonal migrations between low latitude winter breeding grounds and high latitude summer feeding grounds. Although arguably one of the best studied of all baleen whales, there remain some critical gaps in our understanding of their population structure, migratory movement and the mixing of putative populations on the feeding grounds. Addressing these uncertainties is important in the development of demographic models that reconstruct the historical trajectory of population decline and recovery following the cessation of commercial whaling.

Utilising both mitochondrial and nuclear genetic markers, this thesis examines the population structure and distribution of humpback whales that migrate to separate winter breeding grounds along the north-western and north-eastern coasts of Australia, and their interaction with the endangered populations of the South Pacific. The project investigated three important gaps in knowledge: population structure among putative breeding populations, the mixing of breeding populations on high latitude Antarctic feeding grounds and evidence for sex-specific migration along the eastern Australian migratory corridor. The thesis also reports the discovery and utility of novel nuclear genetic markers (single nucleotide polymorphisms, SNPs). These markers hold promise for facilitating more effective multi-laboratory collaboration.

Among the Australian putative populations, weak but significant differentiation was detected across ten microsatellite loci and mitochondrial control region sequences. This pattern of low level differentiation is emerging as a characteristic of Southern Hemisphere humpback whale populations indicating extensive movement at least historically, if not presently.

As the first step towards assessing the mixing of Australian and endangered South Pacific humpback whale breeding populations on the Antarctic feeding grounds, a series of simulations were conducted to estimate the statistical power of both mitochondrial and nuclear microsatellite data from these populations for a mixed-stock analysis (MSA). The results of these simulations confirmed that we can draw robust conclusions from our MSA of Antarctic feeding ground samples collected south of eastern Australia and New Zealand in 2010. Using combined mtDNA and microsatellite datasets revealed substantial contributions from both eastern Australia and New Caledonia, but not western Australia; strengthening emerging evidence that these Antarctic waters are utilized by humpback whales from both eastern Australia and the more vulnerable breeding population of New Caledonia, representing Oceania.

There was no compelling evidence for sex-specific migration within the eastern Australian breeding population as indicated by the lack of significant differences detected in the patterns of haplotype sharing, haplotype frequency or haplotype differentiation between males and females. Instead, the significant differentiation revealed between the sexes at the nucleotide level for one sampling location and between sampling locations at the haplotype level suggests that humpback whale migration along eastern Australia may be more complex than previously thought.

Increasing the statistical power of our genetic datasets through the addition of new informative markers, including the SNPs discovered in this project, and incorporating non-genetic data, will assist in future studies of the population genetic structure and dynamics of Southern Hemisphere humpback whales.

TABLE OF CONTENTS

PREFACE

Two and a half decades have now passed since passage of the International Whaling Commission's (IWC) moratorium on commercial whaling. Prior to this point, over most of the last century, exploitation of large whales was often immense. In the Antarctic alone, more than 2 million whales were killed, and with the revelations of widespread illegal catches by the Soviet Union (Yablokov 1994, Zemsky *et al.* 1995, Zemsky *et al.* 1996), the large Southern Ocean populations were reduced to a fraction of their original size (Clapham *et al.* 1999).

Our knowledge of the present status of the world's baleen whales varies considerably between species, with the recovery of some populations proceeding strongly while others remain highly endangered. Of the eleven species of baleen whale (Order Cetacea, suborder Mysticeti) populations of four species are considered 'critically endangered' based on abundance estimates, including the blue whale, grey whale, bowhead whale and northern right whale. The northern right whale was amongst the first large whales to be hunted on a systematic, commercial basis with intensive shore whaling during the 17^{th} and 18^{th} century and indiscriminate illegal Soviet whaling in the 19^{th} century making this species the most threatened of all baleen whales throughout all of its range (reviewed in Brownell Jr *et al.* 1986).

The humpback whale (*Megaptera novaeangliae*) is arguably the most studied of all the baleen whale species. A coastal species over much of its extensive world-wide range, humpback whales bore the initial brunt of whaling activities in many areas. Often the first species to be taken, it was frequently hunted to commercial extinction, after which other whales were targeted (Tønnessen and Johnsen 1982, Clapham *et al.* 1997).

Hundreds of thousands of humpback whales were killed during the period of commercial whaling throughout much of their range, particularly in the Southern Ocean, with many populations reduced by more than 90% of their original size. As a result, in 1970 the humpback whale was listed as an endangered species under the Endangered Species Conservation Act of 1969 by the United States government. All populations were listed as one global entity under the act as a precautionary measure, regardless of their individual status.

The impact of whaling and the recovery of whale populations is a key focus of the International Whaling Commission (IWC) scientific committee. Estimating the former abundance and maximum environmental carrying capacity of each of these exploited populations and reconstructing the historical trajectory of their decline are essential to accurately assess the true impact of whaling on the marine ecosystem, and to establish a baseline for a population's recovery. This measurement of population recovery plays a crucial role in conservation management schemes, which for the most part, can mean the difference between the recovery or decline of a species.

Standard demographic models rely heavily on historical catch records from commercial whaling as well as estimates of biological parameters such as the rate of reproduction, age at sexual maturity and natural mortality (Best 2001, Clapham 2001, Jackson *et al.* 2008). However, there are a number of uncertainties associated with these models that can substantially over or under-estimate pre-exploitation abundance including discrepancies in taxonomic boundaries, population structure and contemporary abundance.

Modern molecular markers have been useful in helping us address these problems. Specifically, the use of highly variable mitochondrial and nuclear genetic markers have provided high-resolution genetic information, allowing us to distinguish between individuals and populations within species with little phenotypic variation. They also offer clues on aspects of whale migration and movement from high latitude feeding grounds to low latitude tropical breeding grounds.

Utilising both mitochondrial and nuclear genetic markers, this thesis investigated the patterns of population structure among humpback whales that migrate along the east and west coast of Australia, and the neighbouring endangered populations of the South Pacific. Of important relevance to population assessments of Southern Hemisphere humpback whales, the thesis examines population structure among putative breeding populations, the mixing of breeding populations on high latitude Antarctic feeding grounds and evidence for sex-specific migration along the eastern Australian migratory corridor. The thesis also looks at the discovery and utility of novel nuclear genetic markers (single nucleotide polymorphisms, SNPs) that are easier to ascertain, have a well characterised mutation rate and are more universally comparable than the commonly used microsatellite markers.

Field work, which involved the collection of humpback whale skin biopsy samples, was conducted along the migratory corridors of eastern and western Australia and in the Southern Ocean south of eastern Australia and New Zealand as part of a six week Australian-New Zealand Antarctic Whale Expedition (AWE). I was also privileged to have access to an extensive mtDNA and microsatellite database from the breeding grounds of the South Pacific provided by the South Pacific Whale Research Consortium.

This dissertation represents the partial fulfilment of the requirements for the Degree of Doctor of Philosophy (PhD) at the Research School of Biology at the Australian

National University (ANU) and the Australian Marine Mammal Centre (AMMC) at the Australian Antarctic Division.

The outline of the thesis is as follows:

Chapter 1 - Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations.

This chapter evaluated the population genetic structure among Australian humpback whales at both maternally inherited mtDNA and biparentally inherited microsatellite markers, as well as extend previous analyses of mtDNA variation in a comparison of Australian humpback whales and the endangered populations of Oceania.

Chapter 2 - Mixed-stock analysis of humpback whales (*Megaptera novaeangliae*) on Antarctic feeding grounds.

Using a series of simulations, this chapter evaluated the statistical power of microsatellite and mtDNA datasets from the putative humpback whale populations of Australia and Oceania for individual assignment and mixed-stock analysis given available samples size and the patterns of genetic divergence, and 2) Determined ways in which we can improve the accuracy and precision of mixed-stock analyses for these priority populations for future studies. In light of the simulation outcomes, the study then estimated the population composition of Antarctic feeding ground samples collected south of eastern Australia and New Zealand.

Chapter 3 - Re-assessment of the genetic evidence for sex-specific migratory route choice in eastern Australian humpback whales (*Megaptera novaeangliae*).

This chapter investigated evidence for sex-specific migratory behaviour in humpback whales that migrate along the east coast of Australia by assessing the patterns of mitochondrial haplotype sharing, haplotype frequency and haplotype differentiation among males and females.

Chapter 4 - Development and evaluation of single nucleotide polymorphism (SNP) markers for population structure analysis in the humpback whale (*Megaptera novaeangliae*).

In this chapter I developed a suite of informative SNP markers from intron sequences and estimate by computer simulation the statistical power of a combined panel of intronic and exonic SNPs to detect population genetic structure in humpback whales compared with a suite of microsatellite markers.

Appendix – Development and application of SNPs for humpback whale population genetics (literature review).

A review on SNP marker discovery and genotyping techniques as background research to my final chapter on the development and evaluation of SNPs for humpback whales.

LITERATURE CITED

Best, P., Brandao, A. and Butterworth, D.S. 2001. Demographic parameters of southern right whales off South Africa. Journal of Cetacean Research and Management 2:161-169.

Brownell Jr, R.L., P.B. Best and J.H. Prescott, eds. 1986. Right whales: past and present status. Reports of the International Whaling Commission, Special Issue, 10, 289 pp.

Clapham, P., Robbins, J., Brown, M., Wade, P.R. and Findlay, K. 2001. A note on plausible rates on population growth in humpback whales. Appendix 5, Report of the Sub-Committee on the Comprehensive Assessment of Whale Stocks. Journal of Cetacean Research and Management 3:196-197.

Clapham, P.J., S. Leatherwood, I. Szczepaniak and R.L. Brownell. 1997. Catches of humpback and other whales from shore stations at moss landing and trinidad, california, 1919-1926. Marine Mammal Science 13:368-394.

Clapham, P.J., S.B. Young and R.L. Brownell. 1999. Baleen whales: conservation issues and the status of the most endangered populations [Review]. Mammal Review 29:35-60.

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

Tønnessen, J.N. and A.O. Johnsen. 1982. The History of Modern Whaling. C. Hurst and Co., London.

Yablokov, A.V. 1994. Validity of Soviet whaling data. Nature, London 367:108.

Zemsky, V.A., A.A. Berzin, Y.A. Mikhaliev and D.D. Tormosov. 1995. Soviet Antarctic pelagic whaling after WWII: review of actual catch data. Reports of the International Whaling Commission 46:131-135.

Zemsky, V.A., Y.A. Mikhaliev and A.A. Berzin. 1996. Supplementary information about Soviet whaling in the Southern Hemisphere. Reports of the International Whaling Commission 46:131-138.

Chapter 1: Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations

Natalie T. Schmitt[1,6], Michael C. Double[1], Scott Baker[2], Debbie Steel[2], Curt S. Jenner[3], Micheline-Nicole M. Jenner[3], David Paton[4], Rosemary Gales[5], Simon N. Jarman[1], Nick Gales[1], James R. Marthick[1], Andrea M. Polanowski[1], Rod Peakall[6]

[1] Australian Marine Mammal Centre Science, Australian Antarctic Division, Kingston, Tasmania 7050, Australia

[2] Marine Mammal Institute and the Department of Fisheries and Wildlife, Oregon State University, 2030 SE Marine Science Dr, Newport Oregon 97365 USA

[3] Centre for Whale Research (Western Australia), PO Box 1622, Fremantle, Western Australia 6959, Australia

[4] Blue Planet Marine, PO Box 919 Jamison Centre, 2614, ACT Australia

[5] DPIPWE, 134 Macquarie St, Hobart, Tasmania 7000, Australia

[6] Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra ACT 0200, Australia

Correspondence: N. T. Schmitt, Australian Marine Mammal Centre, Australian Antarctic Division, Kingston, Tasmania 7050, Australia;

E-mail: natalie.schmitt@aad.gov.au; natschmitt@hotmail.com

Running title: Genetic differentiation among Australian humpback whales

ABSTRACT

Humpback whales undertake long-distance seasonal migrations between low latitude winter breeding grounds and high latitude summer feeding grounds. We report the first in-depth population genetic study of the humpback whales that migrate to separate winter breeding grounds along the north-western and north-eastern coasts of Australia, but overlap on summer feeding grounds around Antarctica. Weak but significant differentiation between eastern and western Australia was detected across ten microsatellite loci (F_{ST} = 0.005, P = 0.001; D_{EST} = 0.031, P = 0.001, n = 364) and mitochondrial control region sequences (F_{ST} = 0.017 and Φ_{ST} = 0.058, P = 0.001, n = 364). For the microsatellite markers, Bayesian clustering analyses could not resolve any population structure unless sampling location was provided as a prior. This study supports the emerging evidence that weak genetic differentiation is characteristic among Southern Hemisphere humpback whale populations. This may in part reflect the circumpolar distribution of summer feeding grounds that lack continental barriers, allowing for extensive whale movement.

Keywords: mtDNA, microsatellites, population genetic structure, conservation, management, *Megaptera novaeangliae*

INTRODUCTION

For many marine species, ecological and environmental discontinuities such as ocean currents, changes in bathymetry and ocean temperature are increasingly being identified as cryptic barriers to gene flow and dispersal (Kaschner *et al.* 2006, Knutsen *et al.* 2009, Unal and Bucklin 2010, Mikkelsen 2011, Shen *et al.* 2011). The influence of social and learnt behaviors that may also establish or reinforce population boundaries are less understood. Such factors may be highly relevant to cetacean species that exhibit complex communication and social behaviors and where migratory behavior is thought to be learned through through social inheritance from the mother to the calf (Clapham 1996, Hauser *et al.* 2007). Therefore, despite their high vagility, cetaceans may exhibit highly structured populations primarily driven by non-physical barriers (Hoelzel 1998).

Like other balaenopterid species, humpback whales undertake long-distance seasonal migrations between low latitude winter breeding and calving grounds and high latitude summer feeding grounds (Figure 1, Mackintosh 1965). These whales also exhibit a large range of social and sexual behaviors, have strong maternal fidelity, and are renowned for their repertoire of complex culturally acquired ‘songs’ and calls (Clapham 1996, Noad *et al.* 2000, Valsecchi *et al.* 2002, Smith *et al.* 2008). Historically, humpback whale populations have been defined based on the distribution of calving areas and migratory routes. These populations have been treated as management units in the apportionment of catch quotas for commercial whaling (Kellogg 1929, Chittleborough 1965, Mackintosh 1965, Dawbin 1966). More recently, because demographic studies are difficult to undertake, genetic analysis of mitochondrial (mtDNA) and nuclear markers has been applied to gain insights on population structure, dispersal and mating systems.

Much of what we know about population differentiation among humpback whales has resulted from studies in the Northern Hemisphere, where whales are geographically separated by the American and Asia-European continents (Baker *et al.* 1986, Palsbøll *et al.* 1995, Calambokidis *et al.* 1996, Clapham 1996, Palsbøll *et al.* 1997*a*, Clapham *et al.* 1999, Calambokidis *et al.* 2001, Calambokidis *et al.* 2008). Within each ocean basin, individuals show strong fidelity to specific foraging areas and mix on common breeding grounds (Calambokidis *et al.* 2001, Stevick *et al.* 2006).

In contrast to the Northern Hemisphere where foraging areas are numerous and discrete, humpback whales in the Southern Hemisphere have a circumpolar distribution on high latitude feeding grounds in the Southern Ocean. On their annual migration, they segregate onto seven low latitude breeding areas which are widely distributed around oceanic islands and specific coastal regions proximate to continental shelf areas (Mackintosh 1965). With no continental barriers to dispersal on feeding grounds, there is the potential for frequent movement between populations as described for other marine megafauna (Bonfil *et al.* 2005, Boyle *et al.* 2009).

Two recognized populations of humpback whales occur along the coasts of Australia. One migrates along the eastern seaboard and is thought to mate and calve within the Great Barrier Reef, the other migrates along the western seaboard and mates and calves off the Kimberley coast off western Australia (Jenner *et al.* 2001). During the 20th century, Australian humpback whales were hunted along both the eastern and western migratory corridors and intensively in their Antarctic feeding grounds (Mackintosh 1965). By the time commercial whaling ceased in 1963, the western Australian population was estimated to be fewer than 500 animals from approximately 17,000 prior to 1934 (Chittleborough 1965, Bannister 1994), and the eastern Australian population was reduced to as few as 100 individuals from a pre-exploitation abundance estimate of between 16,022 and 22,957 (Chittleborough 1965, Paterson *et al.* 1994, Jackson *et al.* 2008). Recent data have shown that both populations are recovering strongly with the current rate of increase at about 10-11% per annum (Noad *et al.* 2011, Paxton *et al.* 2011, Salgado Kent *et al.* 2012). Absolute abundance for western Australian humpback whales is currently estimated at 21,750 (95% CI 17,550-43,000) (Hedley *et al.* 2011) and 14,522 (95% CI 12,777-16,504) for eastern Australia (Noad *et al.* 2011).

The degree of connectivity between the Australian populations is poorly understood but migration between the populations has been documented. During the 1950s and 1960s stainless steel 'Discovery' marks were shot into whales and some were then recovered when the whales were killed and flensed. This approach provided the first means of tracking the movement of whales over large distances and periods of time (Mackintosh 1965, Dawbin 1966). In the summer of 1958-59, eleven whales marked on the east coast of Australia were recaptured with two recovered in whales migrating along the west coast, indicating some movement between breeding populations (Chittleborough 1961, 1965, Dawbin 1966).

Here, based on extensive sampling, we specifically evaluate (i) the population genetic structure among the eastern and western populations of Australian humpback whales by examining variation in both maternally inherited mtDNA and biparentally inherited microsatellite markers for both sexes, (ii) extend previous analyses of mtDNA variation among humpback whales in Oceania and western Australia by combining our data with Olavarria *et al*. (2007) to include eastern Australia, (iii) compare and contrast our findings with other studies of humpback whales and consider the ecological implications of the emerging genetic patterns.

METHODS

Population definition in the Southern Hemisphere

The International Whaling Commission (IWC) currently recognizes seven breeding aggregations in the Southern Hemisphere as ‘populations’ A to G, with questions remaining about further subdivision around Africa, Australia, and the South Pacific (Chittleborough 1965, Mackintosh 1965) (Fig. 1). The humpback whales that migrate along the west and east coasts of Australia are recognized as population D and subpopulation E1, respectively. Subpopulations E2 and E3 (Tonga and New Caledonia), and F1 and F2 (Cook Islands and French Polynesia) are often referred to in IWC literature as ‘Oceania’ which is listed separately by the IUCN as endangered (IWC 1998, Childerhouse *et al.* 2008).

Sample collection, DNA extraction and sex identification

A total of 364 biopsy samples were collected from humpback whales. These samples were collected from eastern (Eden, New South Wales; eastern Tasmania) and western Australia (Exmouth). The timing and location of the sampling is presented in Table 1. Samples were collected using a biopsy dart propelled by a modified 0.22 caliber rifle and then stored in 70% ethanol at -80˚C (Krützen *et al.* 2002). Total cellular DNA was extracted from skin tissue using a standard salt extraction technique (Aljanabi 1997), or an automated Promega Maxwell ® 16 System. Sex was determined using a fluorescent 5’exonuclease assay producing PCR product from the ZFX and ZFY orthologous gene sequences (Morin *et al.* 2005).

Microsatellite loci

Samples were genotyped at ten polymorphic microsatellite loci including nine dinucleotide repeats [EV1, EV14, EV37, EV94, EV96 (Valsecchi and Amos 1996); GT211, GT23, GT575 (Berube *et al.* 2000); rw4-10 (Waldick *et al.* 1999)] and one tetranucleotide repeat [GATA417 (Palsbøll *et al.* 1997*b*)]. To allow simultaneous amplification of several loci in one PCR reaction, we used a Qiagen Multiplex Kit for the following sets of loci: set 1 (EV37 and GT23); set 2 (EV14, EV96 and GATA417); set 3 (EV1, EV94 and GT575) and GT211 and rw4-10 individually. For each locus, one of the primers within each pair was labelled fluorescently at the 5' end to allow for visualization of alleles on an automated sequencer. Each PCR had a final volume of 12.5μL and included: 1x Qiagen Multiplex PCR Master Mix (containing HotStarTaq®DNA Polymerase, Multiplex PCR Buffer, $MgCl_2$ and dNTP mix), 2μM of each primer (labelled and non-labelled) and 1 to 8ng of template DNA (estimated using a NanoDrop spectrometer 3300). The thermocycling profile consisted of an initial denaturing step of 95˚C for 15 minutes, 30 cycles (30 s at 94˚C, 90 s at 58˚C annealing, 60 s at 72˚C) followed by a final extension step of 30 minutes at 60˚C, with the exception that the optimal annealing temperature for the single locus reactions (GT211 and rw4-10) was 53˚C. The annealing temperature for the single locus reactions was lowered to 53˚C because null alleles were detected when run at 58˚C.

Fluorescently labelled PCR products were resolved on an ABI 3130 automated sequencer. Allele sizes in base pairs (bp) were determined using the LIZ-500 size standard run in each lane. Microsatellite alleles were visualized and scored using GeneMapper v3.7® (Applied Biosystems).

Microsatellite validation

Four steps were taken to ensure a robust microsatellite analysis. 1) To estimate genotyping error rate (Bonin *et al.* 2004) a subset of 16 samples was randomly selected, DNA extracted and genotyped at all ten loci individually by an independent geneticist. 2) Samples with identical matching genotypes across all ten loci were assumed to be due to repeated sampling and were removed from the data set (see Results). The average probability that two unrelated animals share the same genotype by chance alone, PI (probability of identity), and the more conservative probability, PI_{SIBS} (probability of identity siblings), were calculated following Peakall *et al.* (2006). 3) MICROCHECKER version 2.2.3 (van Oosterhout *et al.* 2004, van Oosterhout *et al.* 2006) was used to screen the microsatellite data set for genotyping errors such as null alleles, stuttering and large allele dropout. 4) Using Arlequin 3.1 (Excoffier *et al.* 2005), we tested for deviation from Hardy-Weinberg equilibrium at each locus and for linkage disequilibrium between loci within each population and among populations. Sequential Bonferroni correction was applied to all multiple pairwise comparisons (Rice 1989).

mtDNA

We amplified an approximately 700bp fragment of the control region proximal to the tPro RNA gene via PCR reaction using primers light-strand M13Dlp1.5 and heavy strand Dlp8 (Garrigue *et al.* 2004). Amplifications were conducted in a final volume of 10µL at the following concentrations: 2.5mM $MgCl_2$, 200µM dNTP, 0.4mM each primer, 0.25U Taq (New England BioLabs ®Inc.), 1 X PCR buffer (10mM Tris-HCl, 50mM KCl, 1.5mM $MgCl_2$) and 1µL DNA (approximately 10-50ng). Temperature profiles consisted of an initial denaturing period of 2 minutes at 94˚C, followed by 35 cycles of denaturation at 94˚C for 30 seconds, annealing at 54˚C for 40 seconds, and extension at 72˚C for 40 seconds. A final extension period for 10 minutes at 72˚C was also included. Unincorporated primers were removed from PCR products using ExoSAP-IT® or Agencourt AMPure XP. Sequencing reactions with the PCR primers were run using a Big Dye terminator sequencing kit v3.1 (Applied Biosystems) followed by the use of Agencourt CleanSEQ to remove unincorporated primers. PCR products were sequenced on an ABI 3130 automated sequencer.

Forward and reverse sequences were manually edited, trimmed and aligned within Sequencher®4.8 (Gene Codes Corp.) against sequences of 470bp in length, representing the panel of haplotypes previously defined from the South Pacific (Olavarría et al. 2007). This region started at position six of the reference humpback whale control region sequence (GenBank X72202; see Baker and Medrano-Gonzalez 2002, Olavarria et al. 2007), and is considered to include more than 85% of the variation in the entire control region. Comparisons of sequences to identify polymorphic sites and haplotypes were conducted using GenAlEx 6.3 (Peakall and Smouse 2006).

Statistical analysis

For the purpose of presenting summary statistics, the samples from Eden and Tasmania were pooled and are collectively referred to as eastern Australian samples. For each microsatellite locus, the number of alleles, the number of private alleles, the observed heterozygosity and the expected heterozygosity for each geographic region was calculated using GenAlEx 6.3. Arlequin 3.1 (Excoffier *et al.* 2005) was used to determine standard measures of mtDNA genetic diversity including haplotype frequencies, the number of unique haplotypes, the number of shared haplotypes, haplotype and nucleotide diversity, and the number of sequence polymorphic sites.

Haplotype and nucleotide diversity was calculated according to Nei (1987) and Tajima (1983), respectively.

The extent of genetic differentiation among sampling locations and among the two putative populations was evaluated using Analysis of Molecular Variance (AMOVA) (Excoffier *et al.* 1992) as implemented in GenAlEx 6.3, with statistical testing by random permutation (999 permutations). For microsatellite data, an estimate of F_{ST} (infinite allele model) was calculated as per Weir and Cockerham (1984), Peakall *et al.* (1995) and Michalakis and Excoffier (1996). Recent analyses suggest that these standard measures of differentiation may be poorly suited as estimators of population divergence for data sets in which allelic diversity is high (Hedrick 2005, Jost 2008, Meirmans and Hedrick 2011). Given the high variability of the markers used here, Jost's D_{EST}, an unbiased estimator of divergence, was calculated using a modified version of the R package DEMEtics V0.8.0 (Jueterbock *et al.* 2010), with overall estimates of D_{EST} calculated from individual loci using a harmonic mean approximation and statistical testing by bootstrapping with 1000 permutations. Compared with F_{ST}, D_{EST} partitions diversity based on the effective number of alleles rather than on the expected diversity to give an unbiased estimation of divergence (Jost 2008). For mtDNA data, an AMOVA was performed at both the nucleotide and haplotype level. For these analyses, genetic distance matrices were constructed using individual pairwise differences at all polymorphic nucleotide sites, or haplotype differences among all individuals (Griffiths *et al.* 2011). In keeping with the common practice in similar studies of humpback whales (Olavarría *et al.* 2007, Rosenbaum *et al.* 2009) we use the notation F_{ST} for haplotype differentiation and Φ_{ST} for nucleotide differentiation (e.g. Weir and Cockerham 1984, Takahata and Palumbi 1985, Hudson *et al.* 1992).

To evaluate the genetic data without the need to impose *a priori* population structure, we applied the Bayesian clustering approach implemented in the software STRUCTURE version 2.3.1 (Pritchard *et al.* 2000) to the microsatellite data set. We also repeated the analysis using the three sampling locations as priors to assess the influence of geography (LocPrior model; Hubisz *et al.* 2009). This method attempts to partition samples into K groups such that the loci in those groups are in Hardy-Weinberg equilibrium, and linkage equilibrium. An ancestry model of admixture and correlated allele frequencies were assumed among populations with 10,000 burn-in steps and 300,000 Markov Chain Monte Carlo repetitions. Five replicates for each number of populations (K =1 to 6) were performed to verify that the number of populations identified was consistent between runs. STRUCTURE output was summarized and evaluated using the software *CorrSieve* (Campana *et al.* 2011).

Potential evidence for strong sex-biased dispersal between eastern and western Australia was investigated by calculating pairwise estimates of F_{ST} among populations for each sex separately using an AMOVA in GenAlEx 6.3 for both genetic markers. For comparative purposes, Jost's D_{EST} was also calculated for microsatellite data. D_{EST} was not calculated for mtDNA data as the method is based on differences in interpopulation gene diversity (Jost 2008), and as such, does not take into account the evolutionary relationships between haplotypes (Meirmans and Hedrick 2011).

To investigate genetic structure between the Australian populations and those of the South Pacific (including New Caledonia, Tonga, Cook Islands, French Polynesia and Columbia), we combined our mtDNA data with those presented by Olavarria *et al.* (2007) and calculated F_{ST} and Φ_{ST} for pairwise comparisons. The correlation between geographic and genetic distances was analysed using a Mantel test with statistical testing based on 999 random permutations conducted in GenAlEx 6.3 (Smouse *et al.*

1986, Smouse and Long 1992). Correlation coefficients were calculated between F_{ST} and Φ_{ST}, and the geographic distances between all sampling locations.

RESULTS

Each of the ten microsatellite loci were found to be in Hardy-Weinberg equilibrium (Table 2) and pairwise comparisons between loci revealed no linkage disequilibrium (all values of $P > 0.01$) after sequential Bonferroni correction. MICROCHECKER found no evidence of null alleles or stutter/short allele dominance effects across microsatellite loci, with null allele frequency estimates listed for each region in Table S1, Supplementary information. Repeat genotyping of 16 samples by an independent geneticist revealed two inconsistencies across 320 alleles – an error rate of 0.6%. This rate is lower than suggested by the guidelines of the IWC (IWC 2008) for systematic quality control in the use of microsatellite markers (≤ 10% error rate) for management decisions. This low error rate does not guarantee that these genotypes are in fact correct, but provides a significant increase in probability that they are correct compared to a single genotyping event (Pompanon *et al.* 2005).

Sample size and sex ratio

The 364 samples generated 336 unique microsatellite genotypes suggesting the sample set included 28 duplicate samples (resampling the same individual within a pod) (Table 1), with no matches between sampling locations. After removal of the duplicate genotypes the average probability of identity calculated using all remaining genotyping was 6.8×10^{-14} ($PI_{SIBS} = 3.3 \times 10^{-5}$) as calculated from the formulas shown in Peakall *et al.*(2005). These very low probabilities, and the fact that each of the 28 pairs were also of the same sex and mtDNA haplotype, justify the removal of the putative duplicate samples.

The sex ratio of the overall sample was significantly biased toward males (197 males to 139 females, $\chi^2 = 10.39$, $P < 0.01$) as were the eastern Australian samples separately (81 males to 50 females, $\chi^2 = 7.34$, $P < 0.01$). The sex ratio of the western Australian samples did not differ significantly from parity (116 males to 89 females, $\chi^2 = 3.56$, $P = 0.06$) (Table 1).

Genetic diversity

Summary data for each microsatellite locus are presented in Table 2. Across all ten loci, the mean number of alleles per locus was 11.4 and 11.2 for eastern and western Australia, respectively, ranging from four (EV1) to 19 alleles (EV37). There were 120 alleles in total, eight of which were private to eastern Australia with six private to western Australia. Mean expected heterozygosity across loci was similar for both western and eastern Australia (0.81 ± 0.03 and 0.80 ± 0.03, respectively). Bootstrap resampling of the western Australian data set was conducted to generate ten data sets of the same size as eastern Australia.

Of the 336 samples representing unique genotypes, 289 sequences of 470bps in length were used in all subsequent analyses (104 from eastern Australia and 185 from western Australia); 33 could not be sequenced and 14 samples produced ambiguous base calls within the target sequence. Within these sequences 65 polymorphic sites were identified (two indels, two transversions and 61 transitions) which defined 73 haplotypes (Fig. S1, Supporting information). Of these 73 haplotypes, 40 were found only in western Australia and 17 only in eastern Australia (Table 3). Overall haplotype and nucleotide diversities were 0.98 ± 0.003 and 0.02 ± 0.01, respectively. The haplotype and nucleotide diversity for western and eastern Australia are presented in Table 3. Resampling of the western Australian data set to generate ten data sets of equivalent size to the eastern Australian data set showed similar diversity estimates (haplotype diversity = 0.97 ± 0.01, nucleotide diversity = 0.02 ± 0.01).

Genetic differentiation and population structure analysis

Pairwise comparisons between sampling locations in an AMOVA analysis found no significant differentiation between Eden and Tasmania for either the microsatellite (infinite allele model F_{ST} = -0.0003, P = 0.5; D_{EST} = -0.002, P = 0.6) or the mtDNA (haplotype level F_{ST} = -0.006, P = 0.5; nucleotide level Φ_{ST} = -0.01, P = 0.4) data sets. In contrast, significant differentiation was found between Eden and western Australia, and Tasmania and western Australia (see below). This result, together with the known timing of migration and satellite tracking data (Gales *et al.* 2009), suggests the whales sampled off Eden and Tasmania are likely to be from the same population and were therefore combined in all subsequent analyses to represent the eastern Australian population.

The AMOVA analysis found significant structure between the eastern and western Australian populations for mtDNA at the haplotype and nucleotide level (F_{ST} = 0.017, P = 0.001; Φ_{ST} = 0.058, P = 0.001). For microsatellite data, there was also significant but low differentiation between populations using the infinite allele model of mutation (F_{ST} = 0.005, P = 0.001) and Jost's D_{EST} (D_{EST} = 0.031, P = 0.001).

When the STRUCTURE simulation was run without any priors on the geographic origin of samples, only one population was detected for microsatellite data (*Pr(k)* > 0.99). When the three sampling locations were provided as priors however, the results indicated evidence (highest posterior probability) for two populations consisting of western Australia vs. the two eastern sampling locations combined (average estimated *ln* probability: K = 1: -13270; K = 2: -13250; K = 3: -13677; K = 4: -13503; K = 5: -

13674; K = 6: -13990) (Fig. 2a). This result was confirmed by the *CorrSieve* calculation of ΔK and ΔF_{ST}, with maximum values for both equations at $K = 2$ (Fig. S2).

Pairwise analyses for microsatellite data showed significant structure between the two populations for males ($F_{ST} = 0.007$, $P = 0.001$; $D_{EST} = 0.04$, $P = 0.001$) but not for females for F_{ST} ($F_{ST} = 0.002$, $P = 0.07$) after sequential Bonferroni correction. Significant differentiation however, was detected for females between populations using Jost's D_{EST} ($D_{EST} = 0.02$, $P = 0.01$). In pairwise analyses of mtDNA, both males and females showed significant structure between populations at the haplotype and nucleotide level (females: $F_{ST} = 0.02$, $P = 0.002$; $\Phi_{ST} = 0.08$, $P < 0.001$ and males: $F_{ST} = 0.01$, $P = 0.002$; $\Phi_{ST} = 0.04$, $P < 0.001$ after sequential Bonferroni correction). Due to the low levels of differentiation detected in these analyses, we did not examine whether the difference between F_{ST} male and F_{ST} female could be attributed to chance sampling, as the differences are unlikely to be significant.

After merging the data sets described here with mtDNA data described by Olavarria *et al*. (2007), which had no data from eastern Australia, we found low but significant differentiation between the eastern Australia population and all six breeding populations represented from Oceania at both the haplotype and nucleotide level after sequential Bonferroni correction (Table 4). The Mantel test revealed significant correlation between genetic and geographic distances suggesting a pattern of increasing genetic differentiation with increasing geographic separation (F_{ST}: $R_{XY} = 0.70$, $P = 0.03$; Φ_{ST}: $R_{XY} = 0.74$, $P = 0.04$).

DISCUSSION

Patterns of genetic differentiation among Australian humpback whales

Both nuclear and mtDNA markers revealed low but significant differentiation between the eastern and western Australian humpback populations. This finding was supported by the detection of two populations by the Bayesian clustering analysis for the microsatellite markers when using *a priori* information on sampling location. However, without priors the Bayesian clustering analysis failed to detect population subdivision which, as noted by other studies (*e.g.* Berry *et al.* 2004, Latch *et al.* 2006), is likely to be a consequence of the relative insensitivity of this approach when population differentiation is weak.

High genetic diversity was found within both the eastern and western Australian populations for each marker. Such high diversity may be surprising given the known population bottlenecks, however, industrial whaling in the Southern Hemisphere was intense but relatively brief (approximately four decades) and the rates of recovery for both populations have been rapid (Hedley *et al.* 2011, Noad *et al.* 2011, Paxton *et al.* 2011, Salgado Kent *et al.* 2012, Sremba *et al.* 2012). These factors together with the long generation time and age structure of humpback whales have all contributed to minimizing the loss of genetic diversity (see Baker *et al*. 1993 for a similar conclusion).

As expected, genetic differentiation between the eastern and western Australian humpback populations was stronger for mtDNA than nuclear DNA. Several factors can contribute to this common pattern including the larger effective population size of nuclear genes, differences in the rate and mode of mutation (Palumbi and Baker 1994, Baker *et al.* 1998b), and sex-biased dispersal (Avise 1995, Balloux *et al.* 2000). However, among Australian humpback whale populations there is limited evidence for strong sex-biased dispersal despite the expectation of female philopatry and male-driven gene flow displayed by many migratory marine vertebrates (Greenwood 1983, Pardini *et al.* 2001, Bowen and Karl 2007, Engelhaupt *et al.* 2009), with similar levels of genetic differentiation at both marker types evident between the sexes in this study.

Is low genetic population differentiation a characteristic of Southern Hemisphere humpback whales?

In the Northern Hemisphere there is strong differentiation among feeding areas of the North Pacific based on mtDNA sequences (F_{ST} = 0.18), and the North Atlantic (K_{ST} = 0.04) (Palsbøll *et al.* 1995, Larsen *et al.* 1996, Baker *et al.* 1998b). Strong population differentiation has also been detected among breeding populations within the North Pacific for mtDNA (F_{ST} = 0.11) and nuclear intron alleles (F_{ST} = 0.07), reflecting long-term isolation between the wintering grounds of the Hawaiian archipelago and the coast of Mexico (Baker *et al.* 1998b, Baker and Steel 2010).

In contrast to the Northern Hemisphere, we have shown that differentiation among the Australian populations is weak at both nuclear microsatellites and mtDNA. Similarly, in the South Atlantic the degree of genetic differentiation between breeding populations is also weak at nuclear and mtDNA markers, even when separated by the African continent (Pomilla *et al.* 2005, Rosenbaum *et al.* 2009) (see Fig. 1). Studies of divergence among the breeding populations of the South Pacific have not yet included nuclear markers but mtDNA analysis indicates patterns of weak structure (Olavarría *et al.* 2007) (Fig. 1). Even among distant breeding populations, such as eastern Australia versus Columbia, compared in this study (see Table 4), F_{ST} values are low for mtDNA (F_{ST} ~ 0.06 compared to a mean F_{ST} ~ 0.13 in the North Pacific; Baker *et al.* 1998b, Baker and Steel 2010). Thus, the emerging evidence suggests that humpback whale populations of the Southern Hemisphere are characterized by weak differentiation. This indicates that at least historically, if not presently, there is extensive gene flow among humpback whale populations in the Southern Hemisphere.

Is non-genetic evidence consistent with the genetic evidence for extensive gene flow among Southern Hemisphere humpback whales?

For both males and females there is non-genetic evidence for ongoing and wide-ranging movement between breeding grounds across the Southern Hemisphere. For example, based on fluke matching, Stevick *et al*.(2011) reported the movement of an individual female humpback whale between the breeding grounds off Brazil and Madagascar, representing a distance of nearly 10,000 km. A photo-identification study over a six year period in the South Pacific reported four re-sightings of male humpback whales between eastern Australia and neighbouring breeding grounds (from catalogues of 1248 and 672 individuals respectively) (Garrigue *et al.* 2011). In a preliminary study, a catalogue of fluke images from eastern Australia recently reported a match with a likely male photographed opportunistically off western Australia (Kaufman *et al.* 2011), supporting movement between these breeding populations. This non-genetic evidence when coupled with the low levels of differentiation suggests that whales moving between breeding areas may be interbreeding.

The analysis of humpback whale song also provides another line of non-genetic evidence in support of gene flow among Southern Hemisphere populations. Male humpback whales sing throughout their migration from the feeding grounds to breeding grounds where the song is transmitted culturally among individuals and is thought to be a form of sexual display (Noad *et al.* 2000). All males in a population produce the same song, which changes over time, and all singers maintain the changes (Winn and Winn

1978, Payne *et al.* 1983). The differences in the theme composition of male song between humpback breeding populations within the same ocean basin are found to increase with distance, reflecting at least historical migratory exchange between geographically close populations (Helweg *et al.* 1998).

Humpback whale song from western Australia was found to replace the song of eastern Australia over only three breeding seasons in an analysis of song evolution (Noad *et al.* 2000). Noad *et al.* (2000) suggested that this song evolution is mediated by the movement of a small number of males between populations although it is possible that singing on feeding grounds may also transfer song types between populations without the movement of individual whales (Mattila *et al.* 1987). Nonetheless, this rapid transmission of song and change in theme composition support contact among males of these two populations somewhere in their annual migratory cycle or on the feeding grounds, suggesting a potential for gene flow south of the breeding grounds or even on feeding areas.

It is clear that the non-genetic evidence is consistent with the genetic evidence for long-range movement among Southern Hemisphere humpback populations. However, neither photo-ID or song analysis offer comparable indices, so we cannot speculate on the magnitude of interchange these combined indices represent.

Why are the patterns of genetic variation among humpback whale populations different between the Northern and Southern Hemispheres?

In the Northern Hemisphere, coastal wind-driven and curl-driven upwelling from continental land barriers have resulted in many localized areas of nutrification and biological production (Checkley Jr. and Barth 2009), creating opportunities for the formation of segregated humpback whale feeding areas. These localized feeding areas may have also emerged through the formation of the Panama Land Bridge and as a result of long glacial and interglacial periods in the Arctic (Avise 2000, Hewitt 2000, Hewitt 2004). Long-term preference of males and females to localized feeding grounds combined with natal philopatry may explain the comparatively high levels of genetic differentiation between both breeding and feeding populations. By contrast, in the high latitudes of the Southern Ocean, prey density is high and widely distributed throughout a broad, circumpolar area (Williams *et al.* 2010) where glacial barriers have not fluctuated to the same extent (Barker *et al.* 2009), increasing the potential for long-term mixing and therefore gene flow among breeding populations. The extent to which humpback populations mix on these feeding grounds is therefore more likely to depend merely upon the distance between them (Hoelzel 1998).

Although migratory behavior is thought to be socially-inherited from the mother to her calf (Clapham 1996), the mixing of breeding populations in the Southern Ocean may reduce the degree of natal fidelity, relative to that found in the Northern Hemisphere. Also, although it is expected juveniles rather than adults are more likely to move between populations while on the feeding grounds in the Southern Ocean (Clapham 1996), there is growing evidence of adult movement too. In addition to the Discovery marking and recovery described earlier (Chittleborough 1961, Dawbin 1966), photo-identification of humpback (Garrigue *et al.* 2000, Garrigue *et al.* 2002, Kaufman *et al.* 2011) and other baleen whales (Pirzl *et al.* 2009) have all revealed movement of mature whales between breeding populations.

Estimating gene flow in the Southern Hemisphere?

The very low genetic differentiation that appears to characterize humpback populations of the Southern Hemisphere presents challenges for reliable estimates of the magnitude of gene flow. Allendorf *et al*. (2007) suggest that for reliable estimates of *Nm* based on F_{ST}, the levels of differentiation need to be moderate to large (F_{ST} > 0.05 to 0.10). Furthermore, they warn against interpreting *Nm* values literally at the low F_{ST} values found in this study. Similarly, more complex methods for estimating migration, such as the coalescent- and assignment-based approaches are equally unreliable at low levels of genetic divergence (Faubet *et al.* 2007, Palsbøll *et al.* 2010) such as those that characterize Southern Hemisphere humpback populations. For this reason, along with the fact that the accuracy of coalescent-based methods can also be seriously compromised by errors in sample size and mutation rate estimates (Karl *et al.* 2012), gene flow estimates were not included in the present study.

We suggest that while the emerging evidence for gene flow among populations is compelling, additional development is required before we can quantify the magnitude of gene flow with any degree of certainty. With further development, novel approaches such as the kinship-based analyses of the spatio-temporal distribution of related individuals may be able to yield more reliable estimates of current migration rates even at low levels of differentiation (Palsbøll *et al.* 2010). Alternatively, although requiring considerable time and effort, multi-state capture-recapture models of current movement using photo-identification and/or genotype data may prove to be the most reliable method for quantifying the magnitude of gene flow.

ACKNOWLEDGEMENTS

The first author (NTS) was supported by an Australian National University postgraduate research scholarship. Funding to support the research was provided by the Australian Marine Mammal Centre at the Australian Antarctic Division.

Collection of biopsy samples was conducted under permits from NSW, WA and Tasmania. Animal ethics approval for the research was given by the Animal Experimentation Ethics Committee at the Australian National University, Canberra.

We thank David Donnelly for his invaluable voluntary assistance in the field.

We also thank the Sapphire Coast Marine Discovery Centre for their logistic and institutional support in Eden.

We also acknowledge the contribution of Mathew Oakes and Glenn Jacobson at the Multi-Media Centre at the Australian Antarctic Division for his help in designing Figure 1. The base-map for this figure was provided by David Smith at the Australian Antarctic Data Centre which includes data from the Antarctic Digital Database version 5 ©Scientific Committee on Antarctic Research 1993-2006.

LITERATURE CITED

Aljanabi, S.M., Martinez, I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research 25:4692-4693.

Allendorf, F.W., Luikart, G. 2007. Conservation and The Genetics of Populations. Blackwell Publishing, Oxford, UK.

Avise, J.C. 1995. Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. Conservation Biology 9:686-690.

Avise, J.C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge.

Baker, C.S., L. Florez-Gonzalez, B. Abernethy, H.C. Rosenbaum, R.W. Slade, J. Capella and J.L. Bannister. 1998a. Mitochondrial DNA variation and maternal gene flow among humpback whales of the Southern Hemisphere. Marine Mammal Science 14:721-737.

Baker, C.S., L.M. Herman, A. Perry, W.S. Lawton, J.M. Straley, A.A. Wolman, G.D. Kaufman, et al. 1986. Migratory movement and population structure of humpback whales (*Megaptera novaeangliae*) in the Central and Eastern North Pacific. Marine Ecology Progress Series 31:105-119.

Baker, C.S. and L. Medrano-Gonzalez. 2002. World-wide distribution and diversity of humpback whale mitochondrial DNA lineages. Pages 84-99 *in* C.J. Pfeiffer, ed. Molecular and Cell Biology of Marine Mammals. Krieger Publishing, Melbourne, FL.

Baker, C.S., L. Medranogonzalez, J. Calambokidis, A. Perry, F. Pichler, H. Rosenbaum, J.M. Straley, et al. 1998b. Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. Molecular Ecology 7:695-707.

Baker, C.S., A. Perry, J.L. Bannister, M.T. Weinrich, R.B. Abernethy, J. Calambokidis, J. Lien, et al. 1993. Abundant Mitochondrial DNA Variation and World-Wide Population Structure in Humpback Whales. Proceedings of the National Academy of Sciences of the United States of America 90:8239-8243.

Baker, C.S. and D. Steel. 2010. geneSPLASH: genetic differentiation of 'ecostocks' and 'breeding stocks' in North Pacific humpback whales. Abstract in Symposium on the results of the SPLASH humpback whale study: Final report and recommendations. Presented 11 October 2009, Quebec City, Canada.

Balloux, F., H. Brunner, N. Lugon-Moulin, J. Hausser and J. Goudet. 2000. Microsatellites can be misleading: and empirical and simulation study. Evolution 54:1414-1422.

Bannister, J.L. 1994. Continued increase in Group IV humpbacks off Western Australia. Report for the International Whaling Commission. 44: 309-310.

Barker, S., P. Diz, M.J. Vautravers, J. Pike, G. Knorr, I.R. Hall and W.S. Broecker. 2009. Interhemispheric Atlantic seesaw response during the last deglaciation. Nature 457:1097-U1050.

Berry, O., M.D. Tocher and S.D. Sarre. 2004. Can assignment tests measure dispersal? Molecular Ecology 13:551-561.

Berube, M., H. Jorgensen, R. McEwing and P.J. Palsboll. 2000. Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. Molecular Ecology 9:2181-2183.

Bonfil, R., M. Meyer, M.C. Scholl, R. Johnson, S. O'Brien, H. Oosthuizen, S. Swanson, et al. 2005. Transoceanic migration, spatial dynamics, and population linkages of white sharks. Science 310:100-103.

Bonin, A., E. Bellemain, P.B. Eidesen, F. Pompanon, C. Brochmann and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. Molecular Ecology 13:3261-3273.

Bowen, B.W. and S.A. Karl. 2007. Population genetics and phylogeography of sea turtles. Molecular Ecology 16:4886-4907.

Boyle, M.C., N.N. FitzSimmons, C.J. Limpus, S. Kelez, X. Velez-Zuazo and M. Waycott. 2009. Evidence for transoceanic migrations by loggerhead sea turtles in the southern Pacific Ocean. Proceedings of the Royal Society B-Biological Sciences.

Calambokidis, J., E.A. Falcone, T.J. Quinn, A.M. Burdin, P. Clapham, J.K.B. Ford, C.M. Gabriele, et al. 2008. SPLASH: Structure of Populations, Levels of Abundance and Status of Humpback Whales in the North Pacific. Final report for U.S. Dept of Commerce, Western Administrative Center, Seattle, Washington (unpublished).

Calambokidis, J., G.H. Steiger, J.R. Evenson, K.R. Flynn, K.C. Balcomb, D.E. Claridge, P. Bloedel, et al. 1996. Interchange and isolation of humpback whales in the North Pacific. Marine Mammal Science 12:215-226.

Calambokidis, J., G.H. Steiger, J.M. Straley, L.M. Herman, S. Cerchio, D.R. Salden, J. Urban, et al. 2001. Movements and population structure of humpback whales in the North Pacific. Marine Mammal Science 17:769-794.

Campana, M., H. Hunt, H. Jones and J. White. 2011. *CorrSieve*: software for summarizing and evaluating STRUCTURE output. Molecular Ecology Resources 11:349-352.

Checkley Jr., D.M. and J.A. Barth. 2009. Patterns and processes in the California Current System. Progress in Oceanography 83:49-64.

Childerhouse, S., J. Jackson, C. Baker, N. Gales, P. Clapham and R.L. Brownell Jr. 2008. *Megaptera novaeangliae* (Oceania subpopulation) IUCN 2012. IUCN Red List of Threatened Species. Version 2012.1.

Chittleborough, R.G. 1961. Australian catches of humpback whales. CSIRO Australian Division of Fisheries, Oceanographic Reports, No.34.

Chittleborough, R.G. 1965. Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). Australian Journal of Marine Freshwater Research 16:33-128.

Clapham, P.J. 1996. The social and reproductive biology of humpback whales: An ecological perspective. Mammal Review 26:27-49.

Clapham, P.J., S.B. Young and R.L. Brownell. 1999. Baleen whales: conservation issues and the status of the most endangered populations [Review]. Mammal Review 29:35-60.

Dawbin, W.H. 1966. The seasonal migratory cycle of humpback whales. Pages 145-170 *in* K.Norris, ed. Whales, Dolphins and Porpoises. University of California Press, Berkeley, California.

Engelhaupt, D., A. Rus Hoelzel, C. Nicholson, A. Frantzis, S. Mesnick, S. Gero, H. Whitehead, et al. 2009. Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (*Physeter macrocephalus*). Molecular Ecology 18:4193-4205.

Excoffier, L., G. Larval and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetic data analysis. Evolutionary Bioinformatics Online 1:47-50.

Excoffier, L., P.E. Smouse and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Faubet, P., R.S. Waples and O.E. Gaggiotti. 2007. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. Molecular Ecology 16:1149-1166.

Gales, N., M.C. Double, S. Robinson, K.C.S. Jenner, M.-N.M. Jenner, E. King, J. Gedamke, et al. 2009. Satellite tracking of southbound East Australian humpback whales (*Megaptera novaeangliae*): challenging the feast or famine model for migrating whales. Paper SC/61/SH17 presented at the IWC Scientific Committee, 2009 (unpublished).

Garrigue, C., A. Aguayo, V.L.U. Amante-Helweg, C.S. Baker, S. Caballero, P. Clapham, R. Constantine, et al. 2002. Movements of humpback whales in Oceania, South Pacific. Journal of Cetacean Research and Management 4:255-260.

Garrigue, C., W. Dodemont, D. Steel and C.S. Baker. 2004. Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm low abundance and reproductive autonomy of humpback whales on the wintering grounds of New Caledonia. Marine Ecology-Progress Series 274:251-262.

Garrigue, C., P.H. Forestell, J. Greaves, P. Gill, P. Naessig, N.M. Patenaude and C.S. Baker. 2000. Migratory movements of humpback whales (*Megaptera novaeangliae*) between New Caledonia, East Australia and New Zealand. Journal of Cetacean Research and Management 2:111-115.

Garrigue, C., T. Franklin, R. Constantine, K. Russel, D. Burns, M. Poole, D. Paton, et al. 2011. First assessment of interchange of humpback whales between Oceania and the east coast of Australia. Journal of Cetacean Research and Management 3:269-274.

Greenwood, P.J. 1983. Mating systems and the evolutionary consequences of dispersal. Pages 116-131 *in* I.R. Swingland, Greenwood, P.J., ed. The Ecology of Animal Movement. Clarendon Press, Oxford.

Griffiths, K.E., J.W.H. Trueman, G.R. Brown and R. Peakall. 2011. Molecular genetic analysis and ecological evidence reveals multiple cryptic species among thynnine wasp pollinators of sexually deceptive orchids. Molecular Phylogenetics and Evolution 59:195-205.

Hauser, D.D.W., M.G. Logsdon, E.E. Holmes, G.R. VanBlaricom and R.W. Osborne. 2007. Summer distribution patterns of southern resident killer whales *Orcinus orca*:

core areas and spatial segregation of social groups. Marine Ecology-Progress Series 351:301-310.

Hedley, S.L., J.L. Bannister and R.A. Dunlop. 2011. Abundance estimates of Southern Hemipshere Breeding Stock 'D' humpback whales from aerial and land-based surveys off Shark Bay, Western Australia, 2008. Journal of Cetacean Research and Management (special issue 3):209-221.

Hedrick, P. 2005. A standardized genetic differentiation measure. Evolution 59:1633-1638.

Helweg, D.A., D.H. Cato, P.F. Jenkins, C. Garrigue and R.D. McCauley. 1998. Geographic variation in South Pacific humpback whale songs. Behaviour 135:1-27.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907-913.

Hewitt, G.M. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philosphical Transactions of the Royal Society B-Biological Sciences 359:183-195.

Hoelzel, A.R. 1998. Genetic structure of cetacean populations in sympatry, parapatry, and mixed assemblages: Implications for conservation policy. Journal of Heredity 89:451-458.

Hubisz, M.J., D. Falush, M. Stephens and J.K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322-1332.

Hudson, R.R., D.D. Boos and N. Kaplan. 1992. A statistical test for detecting geographic subdivisions. Molecular Biology and Evolution 9:138-151.

IWC. 1998. Annex G. Report of the Sub-Committee on the Comprehensive Assessment of Southern Hemisphere Humpback Whales 48:170-82. International Whaling Commission.

IWC. 2008. Annex I. Report of the working group on stock definition. Journal of Cetacean Research and Management (Suppl.) 10:225-232.

Jackson, J.A., A. Zerbini, P. Clapham, R. Constantine, C. Garrigue, N. Hauser, M.M. Poole, et al. 2008. Progress on a two-stock catch allocation model for reconstructing population histories of east Australia and Oceania. Paper SC/60/SH14 presented at the IWC Scientific Committee, 2008 (unpublished).

Jenner, K.C.S., M.-N.M. Jenner and K.A. McCabe. 2001. Geographical and temporal movements of humpback whales in Western Australian waters. Australian Petroleum Production & Exploration Association Journal 41:749 - 765.

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

Jueterbock, A., P. Kraemer and G.D.J. Gerlach. 2010. *DEMEtics v0.8.0.* University of Oldenburg, Oldenburg, Germany.

Karl, S.A., R.J. Toonen, W.S. Grant and B.W. Bowen. 2012. Common misconceptions in molecular ecology: echoes of the modern synthesis. Molecular Ecology doi: 10.1111/j.1365-294X.2012.05576.x.

Kaschner, K., R. Watson, A.W. Trites and D. Pauly. 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. Marine Ecology-Progress Series 316:285-310.

Kaufman, G., D. Coughran, J. Allen, N. Bott, D. Burns, C. Burton, C. Castro, et al. 2011. Photographic evidence of interchange between east Australia (BS E-1) and west Australia (BS – D) breeding populations. Paper SC/63/SH11 presented at the IWC Scientific Committee, 2011 (unpublished).

Kellogg, R. 1929. What is known of the migration of some of the whalebone whales. Smithsonian Institution Annual Report pp467-494.

Knutsen, H., P.E. Jorde, H. Sannaes, A.R. Hoelzel, O.A. Bergstad, S. Stefanni, T. Johansen, et al. 2009. Bathymetric barriers promoting genetic structure in the deepwater demersal fish tusk (*Brosme brosme*). Molecular Ecology 18:3151-3162.

Krützen, M., L.M. Barre, L.M. Möller, M.R. Heithaus, C. Simms and W.B. Sherwin. 2002. A biopsy system for small cetaceans: Darting success and wound healing in Tursiops SPP. Marine Mammal Science 18:863-878.

Larsen, A.H., J. Sigurjonsson, N. Oien, G. Vikingsson and P. Palsboll. 1996. Population genetic analysis of nuclear and mitochondrial loci in skin biopsies collected from central and northeastern North Atlantic humpback whales (*Megaptera novaeangliae*): population identity and migratory destinations. Proceedings of the Royal Society of London - Series B: Biological Sciences 263:1611-1618.

Latch, E.K., G. Dharmarajan, J.C. Glaubitz and O.E. Rhodes Jr. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics 7:295-302.

Mackintosh, N.A. 1965. The stock of whales. Fishing News (Books) Ltd., London.

Mattila, D.K., L.N. Guinee and C.A. Mayo. 1987. Humpback Whale Songs on a North-Atlantic Feeding Ground. Journal of Mammalogy 68:880-883.

Meirmans, P.G. and P.W. Hedrick. 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11:5-18.

Michalakis, Y. and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061-1064.

Mikkelsen, P.M. 2011. Speciation in modern marine bivalves (Mollusca: Bivalvia): Insights from the published record. American Malacological Bulletin 29:217-245.

Morin, P.A., A. Nestler, N.T. Rubio-Cisneros, K.M. Robertson and S.L. Mesnick. 2005. Interfamilial characterization of a region of the ZFX and ZFY genes facilitates sex determination in cetaceans and other mammals. Molecular Ecology 14:3275-3286.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, USA.

Noad, M.J., D.H. Cato, M.M. Bryden, M.N. Jenner and K.C.S. Jenner. 2000. Cultural revolution in whale songs. Nature 408:537-537.

Noad, M.J., R.A. Dunlop, D. Paton and D.H. Cato. 2011. Absolute and relative abundance estimates of Australian east coast humpback whales (*Megaptera novaeangliae*). Journal of Cetacean Research and Management (special issue 3):243-252.

Olavarría, C., C.S. Baker, C. Garrigue, M. Poole, N. Hauser, S. Caballero, L. Flórez-González, et al. 2007. Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. Marine Ecology Progress Series 330:257-268.

Palsbøll, P.J., J. Allen, M. Berube, P.J. Clapham, T.P. Feddersen, P.S. Hammond, R.R. Hudson, et al. 1997a. Genetic Tagging of Humpback Whales. Nature 388:767-769.

Palsbøll, P.J., M. Berube, A.H. Larsen and H. Jorgensen. 1997b. Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. Molecular Ecology 6:893-895.

Palsbøll, P.J., P.J. Clapham, D.K. Mattila, F. Larsen, R. Sears, H.R. Siegismund, J. Sigurjonsson, 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, P.J., M.Z. Peery and M. Berube. 2010. Detecting populations in the 'ambiguous' zone: kinship-based estimation of population structure at low genetic divergence. Molecular Ecology Resources 10:797-805.

Palumbi, S.R. and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtdna of humpback whales. Molecular Biology & Evolution 11:426-435.

Pardini, A.T., C.S. Jones, L.R. Noble, B. Kreiser, H. Malcolm, B.D. Bruce, J.D. Stevens, et al. 2001. Sex-biased dispersal of great white sharks - in some respects, these sharks behave more like whales and dolphins than other fish. Nature 412:139-140.

Paterson, R., P. Paterson and D.H. Cato. 1994. The status of humpback whales *(Megaptera novaeangliae)* in East Australia thirty years after whaling. Biological Conservation 70:135-142.

Paxton, C.G.M., S.L. Hedley and J.L. Bannister. 2011. Group IV humpback whales: their status from aerial and land-based surveys off Western Australia, 2005. Journal of Cetacean Research and Management (Special Issue 3):223-234.

Payne, K., P. Tyack and R. Payne. 1983. Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): a detailed analysis of two seasons in Hawaii. Pages 9-57 *in* R. Payne, ed. Communication and Behaviour of Whales. Boulder: Westview Press.

Peakall, R., D. Ebert, R. Cunningham and D. Lindenmayer. 2005. Mark-recapture by genetic tagging reveals restricted movements by bush rats (*Rattus fuscipes*) in a fragmented landscape. Journal of Zoology 268:207-216.

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

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Peakall, R., P.E. Smouse and D.R. Huff. 1995. Evolutionary implications of allozyme and RAPD Variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. Molecular Ecology 4:135-147.

Pirzl, R., N.J. Patenaude, S. Burnell and J. Bannister. 2009. Movements of southern right whales (*Eubalaena australis*) between Australian and subantarctic New Zealand populations. Marine Mammal Science 25:455-461.

Pomilla, C., P.B. Best, K.P. Findlay, T. Collins, M.H. Engel, G. Minton, P. Ersts, et al. 2005. Population structure and sex-biased gene flow in humpback whales from wintering Regions A, B, C and X based on nuclear microsatellite variation. Paper SC/57/SH13 presented at the IWC Scientific Committee, 2005 (unpublished).

Pompanon, F., A. Bonin, E. Bellemain and P. Taberlet. 2005. Genotyping errors: Causes, consequences and solutions. Nature Reviews Genetics 6:847-859.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Rice, W.R. 1989. Analyzing tables of statistical tests. Evolution 43:223-225.

Rosenbaum, H.C., C. Pomilla, M. Mendez, M.S. Leslie, P.B. Best, K.P. Findlay, G. Minton, et al. 2009. Population Structure of Humpback Whales from Their Breeding Grounds in the South Atlantic and Indian Oceans. PLoS ONE 4:e7318.

Salgado Kent, C., C. Jenner, M. Jenner, P. Bouchet and E. Rexstad. 2012. Southern Hemisphere Breeding Stock 'D' humpback whale population estimates from North West cape, Western Australia. Journal of Cetacean Research and Management 12(1):29-38.

Shen, K.N., B.W. Jamandre, C.C. Hsu, W.N. Tzeng and J.D. Durand. 2011. Plio-Pleistocene sea level and temperature fluctuations in the northwestern Pacific promoted speciation in the globally-distributed flathead mullet *Mugil cephalus*. Bmc Evolutionary Biology 11:17.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.

Smith, J.N., A.W. Goldizen, R.A. Dunlop and M.J. Noad. 2008. Songs of male humpback whales, *Megaptera novaeangliae*, are involved in intersexual interactions. Animal Behaviour 76:467-477.

Smouse, P.E. and J.C. Long. 1992. Matrix correlation analysis in anthropology and genetics. Yearbook Physical Anthropology 35:187-213.

Smouse, P.E., J.C. Long and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology 35:627-632.

Sremba, A.L., B. Hancock-Hanser, T.A. Branch, R.L. LeDuc and C.S. Baker. 2012. Circumpolar diversity and geographic differentiation of mtDNA in the critically endangered Antarctic blue whale (*Balaenoptera musculus intermedia*). PLoS ONE 7:e32579. doi:32510.31371/journal.pone.0032579.

Stevick, P.T., J. Allen, P.J. Clapham, S.K. Katona, F. Larsen, J. Lien, D.K. Mattila, et al. 2006. Population spatial structuring on the feeding grounds in North Atlantic humpback whales (*Megaptera novaeangliae*). Journal of Zoology 270:244-255.

Stevick, P.T., M.C. Neves, F. Johansen, M.H. Engel, J. Allen, M.C.C. Marcondes and C. Carlson. 2011. A quarter of a world away: female humpback whale moves 10 000 km between breeding areas. Biology Letters 7:299-302.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460.

Takahata, N. and S.R. Palumbi. 1985. Extranuclear differentiation and gene flow in the finite island model. Genetics 109:441-457.

Unal, E. and A. Bucklin. 2010. Basin-scale population genetic structure of the planktonic copepod *Calanus finmarchicus* in the North Atlantic Ocean. Progress in Oceanography 87:175-185.

Valsecchi, E. and W. Amos. 1996. Microsatellite markers for the study of cetacean populations. Molecular Ecology 5:151-156.

Valsecchi, E., P. Hale, P. Corkeron and W. Amos. 2002. Social structure in migrating humpback whales (*Megaptera novaeangliae*). Molecular Ecology 11:507-518.

van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538.

van Oosterhout, C., D. Weetman and W.F. Hutchinson. 2006. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. Molecular Ecology Notes 6:255-256.

Waldick, R.C., M.W. Brown and B.N. White. 1999. Characterization and isolation of microsatellite loci from the endangered North Atlantic right whale. Molecular Ecology 8:1763-1765.

Weir, B.S. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Williams, G.D., S. Nicol, S. Aoki, A.J.S. Meijers, N.L. Bindoff, Y. Iijima, S.J. Marsland, et al. 2010. Surface oceanography of BROKE-West, along the Antarctic margin of the south-west Indian Ocean (30-80 degrees E). Deep-Sea Research Part Ii-Topical Studies in Oceanography 57:738-757.

Winn, H.E. and L.K. Winn. 1978. The song of the humpback whale *Megaptera novaeangliae* in the West Indies. Marine Biology 47:97-114.

TABLES

Table 1. Samples collected from individual (N) humpback whales from three locations off the east and west coast of Australia. The number of duplicate samples is also shown, and the number of known female (F) and male (M) individuals.

Region **Sampling site**	**Sampling period**	**Samples**	**No. of duplicates**	***N***	**F**	**M**
eastern Australia		**141**	**10**	**131**	**50**	**81**
Eden	2008	*63*	*2*	*61*	*14*	*47*
	June	45	2	43	8	35
	Oct, Nov	18	0	18	6	12
Tasmania	2006-2008	*78*	*8*	*70*	*36*	*34*
	July	1	0	1	0	1
	Nov, Dec	77	8	69	36	33
western Australia		**223**	**18**	**205**	**89**	**116**
Exmouth	2007 Sept, Oct					
total		**364**	**28**	**336**	**139**	**197**

Table 2. Genetic diversity in humpback whales from eastern and western Australia genotyped at ten microsatellite loci. *N* = number of genotyped individuals per locus and HW = deviation from Hardy-Weinberg equilibrium (p-value); significant at $P < 0.05$ after adjustment for multiple comparison with the sequential Bonferroni test (Rice 1989) (Standard errors in parentheses).

	N		Number of alleles		Number of private alleles		Observed heterozygosity		Expected heterozygosity		HW (p-value)	
Locus	East	West	East	West	East	West	East	West	East	West	East	West
EV14	131	203	9	8	1	0	0.725	0.754	0.748	0.778	0.747	0.459
EV37	131	202	19	19	2	2	0.916	0.931	0.913	0.904	0.364	0.316
EV96	131	202	13	12	1	0	0.863	0.876	0.848	0.869	0.872	0.682
GATA417	131	203	15	15	2	2	0.870	0.911	0.890	0.903	0.622	0.751
GT211	130	203	10	10	0	0	0.785	0.803	0.820	0.836	0.685	0.030
GT23	131	204	9	9	0	0	0.763	0.838	0.797	0.821	0.192	0.621
rw4-10	131	203	12	12	1	1	0.786	0.877	0.831	0.854	0.567	0.809
EV1	130	203	4	4	0	0	0.523	0.567	0.552	0.526	0.429	0.427
EV94	130	202	9	9	0	0	0.792	0.827	0.807	0.809	0.769	0.352
GT575	130	203	14	14	1	1	0.815	0.788	0.811	0.803	0.801	0.260
All loci	130.6 (*0.2*)	202.8 (*0.2*)	11.4 (*1.3*)	11.2 (*1.3*)	8	6	0.784 (*0.034*)	0.817 (*0.033*)	0.802 (*0.031*)	0.810 (*0.034*)	0.605 (*0.069*)	0.471 (*0.077*)

Table 3. Variability in the mtDNA control region of humpback whales sampled along the east and west coasts of Australia (h = haplotype diversity and π = nucleotide diversity), N = number of samples used in analyses.

Region	**N**	**No. of haplotypes**	**No. of unique haplotypes**	**No. of shared haplotypes**	**h ± SD**	**π ± SD**
East	104	33	17	16	0.961 ± 0.006	0.018 ± 0.009
West	185	56	40	16	0.972 ± 0.004	0.019 ± 0.010
total	289	73			0.975 ± 0.003	0.019 ± 0.010

Table 4. Pairwise comparisons (F_{ST} and Φ_{ST}) among the Australian populations and those of the South Pacific with respective stock definitions in parentheses. Data combines mtDNA control region sequences trimmed to a 470bp consensus region from the present study with those of Olavarria *et al.* (2007). P-values based on statistical testing of 999 random permutations were all significant at $P < 0.005$.

WA = western Australia; EA = eastern Australia; NC = New Caledonia; TG = Tonga; CI = Cook Islands; FP = French Polynesia; COL = Columbia.

a) F_{ST}

Region/Stock	WA (D)	EA (E1)
WA (D)		
EA (E1)	0.014	
NC (E2)	0.015	0.012
TG (E3)	0.017	0.011
CI (F1)	0.028	0.031
FP (F2)	0.040	0.044
COL (G)	0.057	0.063

b) Φ_{ST}

Region/Stock	WA (D)	EA (E1)
WA (D)		
EA (E1)	0.032	
NC (E2)	0.015	0.024
TG (E3)	0.018	0.013
CI (F1)	0.023	0.027
FP (F2)	0.043	0.046
COL (G)	0.049	0.061

Table S1. MICROCHECKER (van Oosterhout *et al*. 2004) null frequencies for all microsatellite loci by population.

Locus	Null Present	Null Frequency	
		East	West
EV14	no	0.0208	0.008
EV37	no	-0.0008	-0.0143
EV96	no	-0.0072	-0.0054
GATA417	no	0.0117	-0.0041
GT211	no	0.0216	0.02
GT23	no	0.0232	-0.0124
rw4-10	no	0.0283	-0.0137
EV1	no	0.032	-0.0467
EV94	no	0.0094	-0.0132
GT575	no	-0.0067	0.009

FIGURES

Figure 1. Southern Hemisphere humpback whale population structure and geographic distribution for both nuclear and mitochondrial DNA (mtDNA) markers.

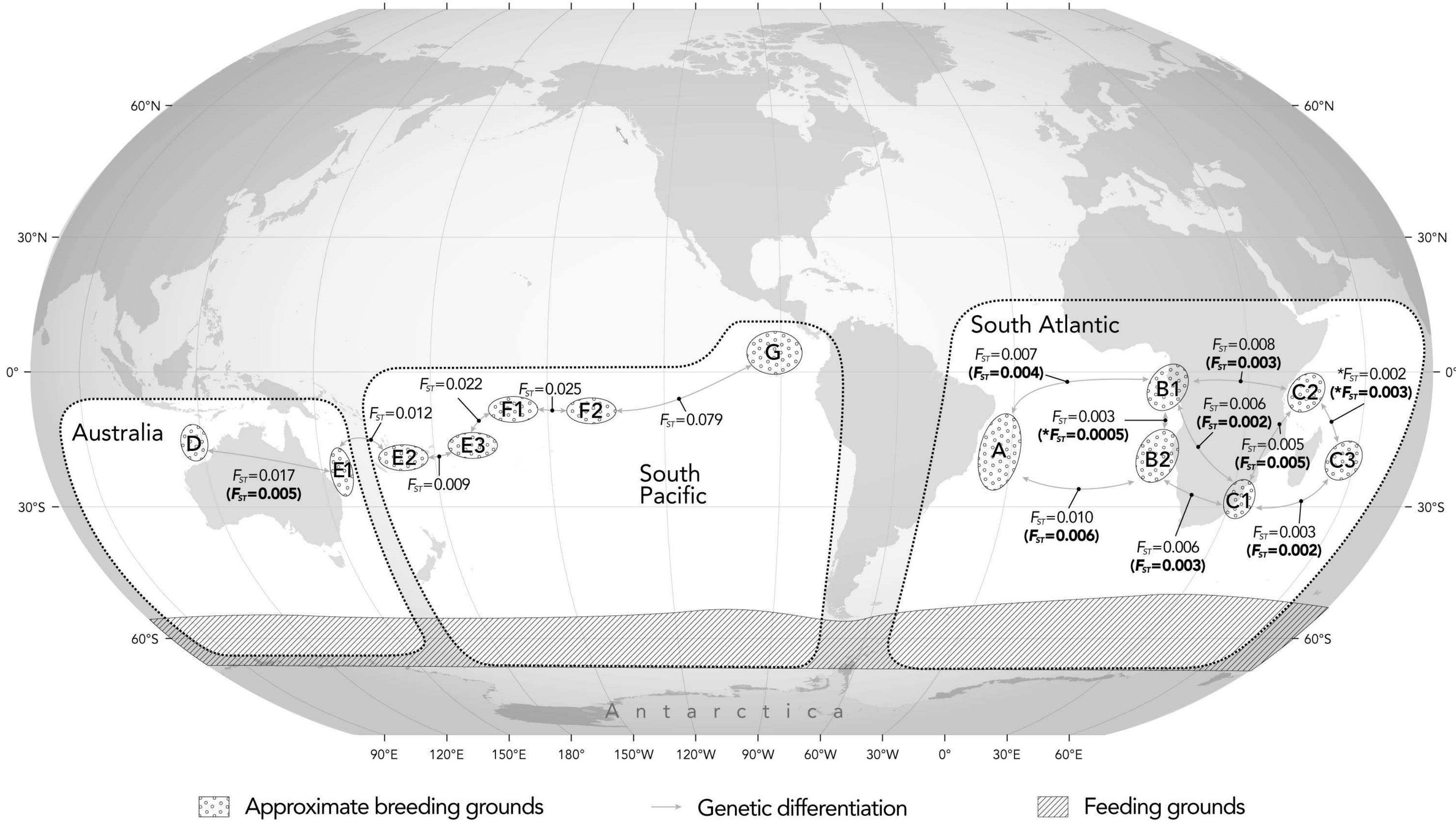
(F_{ST}) **–** Pairwise genetic distances between microsatellite alleles (Australia = ten loci, South Atlantic = nine loci) (Pomilla *et al.* 2005); F_{ST} – Pairwise genetic differences between mtDNA haplotypes (Australia and South Pacific = 470bp, South Atlantic = 486bp) (Baker *et al.* 1998a, Olavarría *et al.* 2007, Rosenbaum *et al.* 2009). All *P*-values less than 0.05 except for *.

Figure 2. Proportional assignment of individual genotypes to each of the K = 2 inferred clusters in the STRUCTURE admixture analysis. Black and grey bars represent proportions of membership to the eastern Australian and western Australian clusters, respectively, using the three sampling locations as priors.

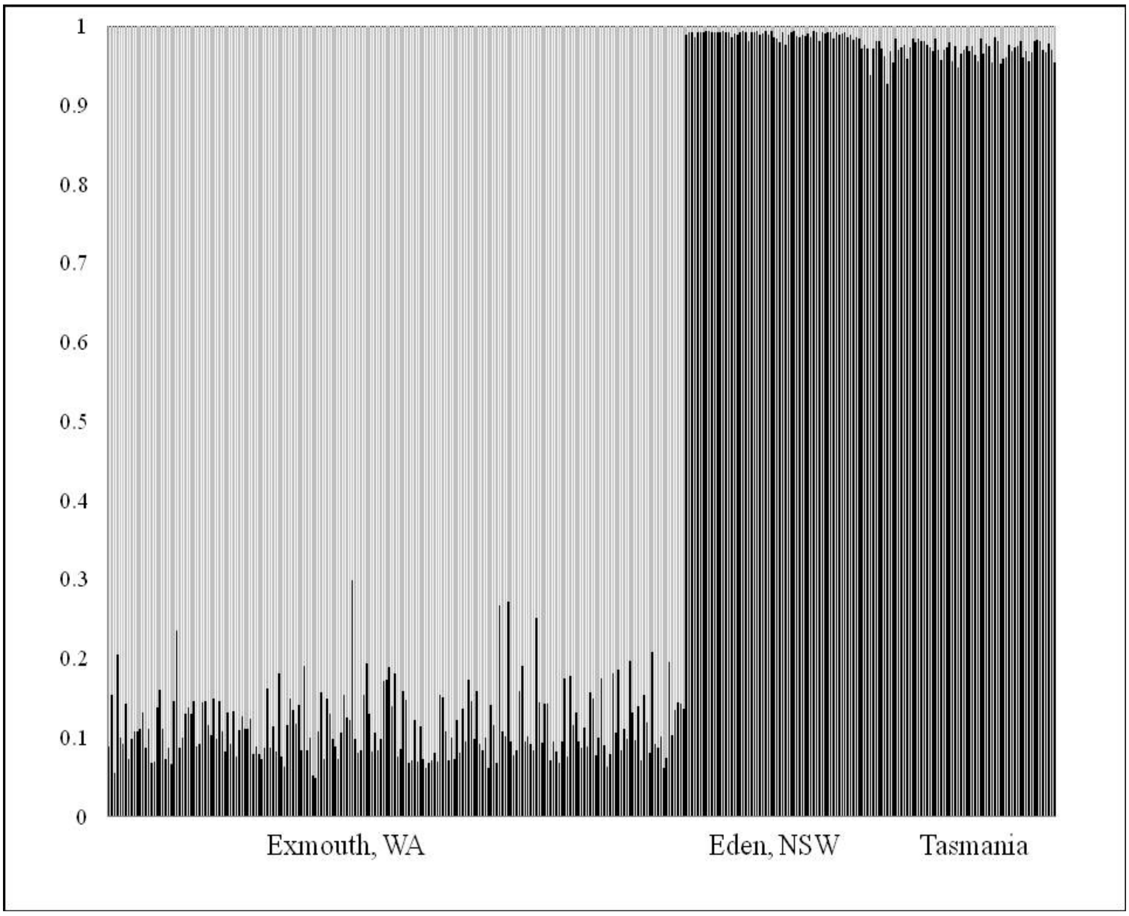
Figure S1. Geographic disribution and relative position of variable nucleotides in humpback whale mtDNA control region defining 73 haplotypes. Dots (.) indicate matches with published reference sequence X72202 (Genbank), dashes indicate insertion/deletion events (Position 1 of alignment corresponds with position 6 of the reference sequence). The total number of each haplotype is indicated for both regions.

* For consistency with the South Pacific haploytpe data set, this polymorphic site was not included in the genetic analyses however, including this locus does not change the number of haplotypes.

Figure S2. Delta K and delta F_{ST} values (ΔK and ΔF_{ST}) calculated using *CorrSieve* for each of the K inferred clusters in STRUCTURE, with a maximum value achieved at K = 2.

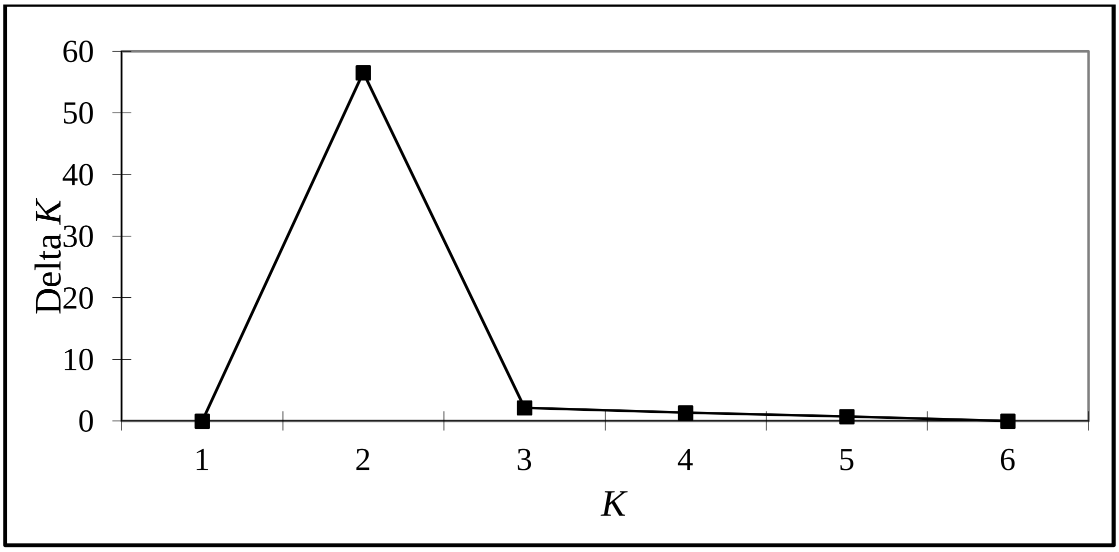


South Atlantic
Australia
South Pacific
Antarctica
A
B1
B2
C1
C2
C3
D
E1
E2
E3
F1
F2
G
F_{ST}=0.007
(F_{ST}=0.004)
F_{ST}=0.008
(F_{ST}=0.003)
*F_{ST}=0.002
(*F_{ST}=0.003)
F_{ST}=0.006
(F_{ST}=0.002)
F_{ST}=0.003
(*F_{ST}=0.0005)
F_{ST}=0.005
(F_{ST}=0.005)
F_{ST}=0.010
(F_{ST}=0.006)
F_{ST}=0.006
(F_{ST}=0.003)
F_{ST}=0.003
(F_{ST}=0.002)
F_{ST}=0.022
F_{ST}=0.025
F_{ST}=0.012
F_{ST}=0.079
F_{ST}=0.009
F_{ST}=0.017
(F_{ST}=0.005)
60°N
30°N
0°
30°S
60°S
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
Approximate breeding grounds
Genetic differentiation
Feeding grounds

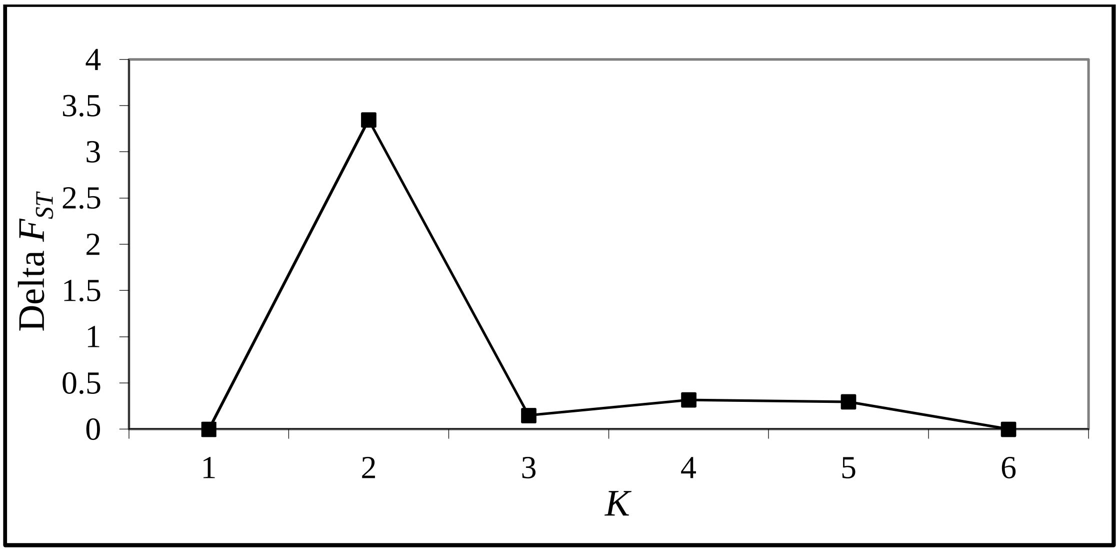

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Exmouth, WA
Eden, NSW
Tasmania

Haplotype	Region East	Region West	Total	8	24	55	57	62	68	75	83	84	92	97	99	101	114	116	117	121	123	126	127	128	134	135	146	161	162	167
ref				T	G	G	T	C	C	T	T	T	G	T	C	A	C	G	-	-	C	C	T	T	A	G	C	C	T	T
1	4		4	C	.	.	.	T	.	.	.	C	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
2	1		1	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	T	.	.
3		5	5	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	T	.	.
4		2	2	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	.	T	.	C
5	2		2	.	.	A	.	T	T	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	T	.	C
6	1		1	.	.	A	.	T	T	.	.	.	.	.	T	.	.	.	-	-	.	.	.	.	.	.	T	T	.	C
7		1	1	.	.	.	C	T	.	C	.	.	.	.	A	.	.	.	-	-	.	T	.	.	.	.	.	T	C	C
8		1	1	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
9		3	3	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
10	1		1	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
11	3	1	4	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	-	-	.	.	.	.	G	.	T	T	.	.
12		3	3	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
13		18	18	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
14		3	3	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
15		3	3	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
16		1	1	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
17	1		1	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
18		1	1	.	.	.	.	.	.	.	.	C	.	.	A	G	.	.	-	-	.	T	.	.	.	.	.	T	C	C
19		1	1	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	-	.	T	.	.	.	.	T	T	.	C
20		1	1	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	C	.	.	.	T	.	.
21		1	1	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
22	3	7	10	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
23		3	3	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
24	4		4	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
25	3	1	4	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
26		3	3	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	.	T	.	.
27		2	2	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	.	.	.	T	T	.	.
28	3		3	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	T	.	.	.	.	G	.	T	T	.	.
29		3	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.
30		4	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	T	.	.
31		3	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	T	.	.
32		2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	G	.	T	T	.	.
33		1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	G	.	T	T	.	.
34		1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	G	.	T	T	.	.
35		1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	G	.	T	T	.	.
36	1		1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
37	1	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
38		8	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
39		5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	G	.	T	T	.	.
40	4	7	11	.	.	.	.	T	.	C	.	.	.	.	A	.	.	.	-	-	.	T	.	.	.	.	.	T	C	C
41		2	2	.	.	.	.	T	.	.	C	C	.	.	.	G	.	.	-	-	.	.	.	.	.	.	T	T	.	C
42		3	3	.	.	.	.	T	.	.	C	.	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	C	C
43		1	1	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	T	.	C
44	1		1	.	.	.	.	T	.	.	.	C	A	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
45	3		3	.	.	.	.	T	.	.	.	C	.	.	A	.	.	A	-	-	.	T	.	.	.	A	.	.	.	C
46		5	5	.	.	.	.	T	.	.	.	C	.	.	A	.	.	A	-	-	.	T	.	.	.	A	.	T	.	C
47	2	2	4	.	.	.	.	T	.	.	.	C	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
48	1	2	3	.	.	.	.	T	.	.	.	C	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
49		6	6	.	.	.	.	T	.	.	.	C	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
50	5	3	8	.	.	.	.	T	.	.	.	C	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
51		1	1	.	.	.	.	T	.	.	.	C	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
52	11	11	22	.	.	.	.	T	.	.	.	C	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
53		1	1	.	.	.	.	T	.	.	.	C	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
54	6	4	10	.	.	.	.	T	.	.	.	C	.	.	A	.	.	.	-	-	.	T	.	.	.	.	.	T	C	C
55	1	2	3	.	.	.	.	T	.	.	.	C	.	.	A	G	.	.	-	-	.	T	.	.	.	.	.	T	C	C
56		3	3	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	T	-	.	T	C	.	.	.	T	T	.	C
57		11	11	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
58	3		3	.	.	.	.	T	.	.	.	C	.	.	.	.	.	.	-	-	T	T	.	.	.	.	.	T	.	C
59	3		3	.	.	.	.	T	.	.	.	C	.	.	T	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
60	4		4	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	-	-	.	T	.	.	.	A	.	T	C	C
61	7	1	8	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	-	-	.	T	.	.	.	.	.	T	C	C
62		2	2	.	.	.	.	T	.	.	.	.	.	.	A	.	T	.	-	-	.	T	.	.	.	A	.	T	C	C
63	1		1	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
64	1	5	6	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	A	.	T	.	C
65	1		1	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	.	T	.	C
66	5	7	12	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
67	6	3	9	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
68	2		2	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
69	9	5	14	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
70		1	1	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
71		3	3	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
72		1	1	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	C	C
73		3	3	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	-	.	T	.	.	.	.	T	T	.	C

Haplotype	171	173	183	189	196	237	240	245	246	247	248	251	253	254	257	261	263	264	265	266	268	269	270	273	284	285	286	306	316	317	347	380	384	437	447	448	469	492*
ref	A	C	T	T	T	T	T	G	T	C	C	A	A	A	T	A	G	T	T	T	T	T	T	G	C	C	T	A	C	T	T	C	C	G	T	T	T	-
1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	C	.	.	.	.	.	.	G
3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
4	.	.	.	.	C	.	C	.	.	T	.	.	.	.	.	.	A	.	C	C	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	G
5	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	C	.	.	A	T	.	.	.	.	.	.	T	.	.	.	C	.	G
6	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	C	.	.	A	T	.	.	.	.	.	.	T	.	.	.	C	.	G
7	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
8	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	C	.	.	.	.	.	.	.	.	.	C	G
9	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	C	.	.	.	.	.	.	.	.	.	.	G
10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
11	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	G
12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	C	.	.	.	C	.	.	.	.	.	.	G
13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	C	.	.	.	.	.	.	G
14	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	C	.	.	.	.	.	.	G
15	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	C	.	.	.	.	.	.	G
16	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	C	.	.	.	.	.	.	G
17	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	C	.	.	.	.	.	.	G
18	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
19	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	.	C	.	.	C	.	.	.	.	C	.	.	G
20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
21	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	C	G
22	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
23	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	G
24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	G
25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
26	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.	.	.	.	.	.	.	.	.	G
27	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
28	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
29	.	.	.	.	.	.	.	.	C	.	T	.	.	.	.	G	.	.	.	.	.	.	.	A	.	T	.	.	.	.	.	.	.	.	.	.	.	G
30	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
31	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G
32	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	A	.	.	.	.	.	.	A	T	.	.	.	.	.	.	T	.	.	.	.	.	G
33	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	G
34	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	G
35	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	G
36	.	.	.	.	.	.	.	.	C	.	.	.	.	.	C	.	.	C	C	C	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
37	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	C	.	.	.	.	.	.	G
38	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	C	.	.	.	.	.	.	G
39	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	C	.	.	.	.	.	.	G
40	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
41	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	C	.	.	A	.	.	.	.	.	.	.	.	.	A	.	C	.	G
42	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	G
43	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	C	.	C	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	G
44	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
45	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	A
46	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	T	.	.	.	.	.	.	.	.	G
47	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
48	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
49	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	G
50	.	.	.	.	.	.	.	.	C	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
51	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	C	.	.	G
52	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
53	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	T	.	.	.	.	.	.	.	.	G
54	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
55	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
56	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	.	C	.	.	.	.	.	.	.	C	.	.	G
57	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
58	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	A	.	C	C	.	.	C	.	T	.	.	.	.	.	.	.	T	.	.	.	.	G
59	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
60	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	C	.	.	C	C	.	.	.	.	.	.	G
61	.	.	.	.	.	.	.	.	C	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
62	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	C	.	.	.	C	.	.	.	.	.	.	G
63	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A	T	.	C	.	.	.	.	.	.	.	C	.	.	G
64	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	.	C	G	.	.	.	.	.	.	C	.	.	G
65	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	A	.	.	C	.	.	.	.	T	.	.	.	.	.	.	T	.	A	.	.	.	G
66	.	.	C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	C	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
67	.	.	.	.	.	.	.	.	C	.	.	.	.	.	C	.	.	C	C	C	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
68	.	.	.	.	.	.	.	.	C	.	.	.	.	.	C	.	.	.	C	C	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
69	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
70	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	G
71	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	G
72	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G
73	.	.	C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	C	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G



60
50
40
30
20
10
0
Delta K
1
2
3
4
5
6
K



4
3.5
3
2.5
2
1.5
1
0.5
0
Delta F_{ST}
1
2
3
4
5
6
K

Chapter 2: Mixed-stock analysis of humpback whales (*Megaptera novaeangliae*) on Antarctic feeding grounds

Natalie T. Schmitt[1,12], Michael C. Double[1], Scott Baker[2,8], Nick Gales[1], Simon Childerhouse[1,8], Andrea M. Polanowski[1,3], Debbie Steel[2], Renee Albertson[2], Carlos Olavarría[4], Claire Garrigue[5,8], Michael Poole[6,8], Nan Hauser[7,8], Rochelle Constantine[8,9], David Paton[8,10], Curt S. Jenner[11], Simon N. Jarman[3], Rod Peakall[12]

[1] Australian Marine Mammal Centre Science, Australian Antarctic Division, Kingston, Tasmania 7050, Australia

[2] Marine Mammal Institute and the Department of Fisheries and Wildlife, Oregon State University, 2030 SE Marine Science Dr, Newport Oregon 97365 USA

[3] Australian Antarctic Division, Kingston, Tasmania 7050, Australia

[4] School of Biological Sciences, University of Auckland, Private bag 92019, Auckland, New Zealand

[5] Opération Cétacés, BP 12827, 98802 Nouméa, Nouvelle-Calédonie

[6] Marine Mammal Research Program, BP 698, 98728 Maharepa, Moorea, French Polynesia

[7] Cook Islands Whale Research, P.O Box 3069, Avarua, Rarotonga, Cook Islands

[8] South Pacific Whale Research Consortium, P.O Box 3069, Avarua, Rarotonga, Cook Islands

[9] The Oceania Project, P.O Box 646, Byron Bay, NSW 2481, Australia

[10] Centre for Whale Research (Western Australia), PO Box 1622, Fremantle, Western Australia 6959, Australia

[11] Blue Planet Marine, P.O Box 919 Jamison Centre, 2614, ACT Australia

[12] Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra ACT 0200, Australia

ABSTRACT

In understanding the impact of commercial whaling, an important gap in knowledge is estimating the mixing of low latitude breeding populations on Antarctic feeding grounds, particularly the endangered humpback whale populations of Oceania. Here we estimated the degree of genetic differentiation among the putative populations of Oceania (New Caledonia, Tonga, the Cook Islands and French Polynesia) and Australia (western australia and eastern Australia) using ten microsatellite loci and mtDNA, 2) assessed the power of the data for individual assignment and mixed-stock analysis (MSA), 3) determined ways we could improve the statistical power of our data for MSA for future studies, and 4) estimated the population composition of Antarctic samples collected in 2010 south of New Zealand and eastern Australia. A large proportion of individuals could not be assigned to a population of origin (>52%) using a posterior probability threshold of > 0.90. The MSA simulations however, produced accurate results with humpback whales reapportioned to their population of origin above the 90% threshold for western Australia, New Caledonia and Oceania grouped using a combined mtDNA and microsatellite dataset. Removing the Cook Islands, considered a transient region for humpback whales, from the simulation analysis increased the ability to reapportion Tonga from 86% to 89% and French Polynesia from 89% to 92%. Breeding ground sample size was found to be a key factor influencing the accuracy of population reapportionment whereas increasing the mixture or feeding ground sample size improved the precision of results. The MSA of our Antarctic samples revealed substantial contributions from both eastern Australia (53.2%, 6.8% SE) and New Caledonia (43.7%, 5.5% SE) [with Oceania contributing 46.8% (5.9% SE)] but not western Australia. Despite the need for more samples to improve estimates of

population allocation, our study strengthens the emerging genetic and non-genetic evidence that Antarctic waters south of New Zealand and eastern Australia are utilized by humpback whales from both eastern Australia and the more vulnerable breeding population of New Caledonia, representing Oceania.

Keywords: humpback whale; population genetics; mixed stock analysis; microsatellite DNA; mtDNA

INTRODUCTION

In conservation and resource management there is often a requirement to assess how breeding populations are impacted either through deliberate or accidental removals. Management tends to focus on breeding units but for migratory species, exploitation or mortality often occurs in other parts of the range. In cases where only a single population is impacted this can be relatively straight-forward but can become complicated when removals occur where populations mix. The assessment then requires an understanding of the degree of mixing and the relative impact on each population contributing to a mix, referred to as mixed-stock analysis (MSA). The classic example of this is the impact of pelagic fishing on salmon populations that exhibit natal philopatry but following smoltification return to the ocean where the mixing of genetically distinct stocks occurs simultaneously with commercial exploitation (e.g. Utter and Ryman 1993, Waples *et al.* 1993, Olsen *et al.* 2000, Cadrin *et al.* 2005, Beacham *et al.* 2008, Beacham *et al.* 2011). Under such conditions, the development of a mixed stock analysis model has been useful in minimizing the risk of overexploiting less productive stocks in the mixed stock fishery.

During the era of industrial whaling in the southern hemisphere, >2 000 000 whales were killed, driving some populations to near extinction (Clapham and Baker 2002). The waters of Antarctica were heavily targeted because many baleen whale populations migrate to and mix in these krill-rich high-latitude waters during the summer. They subsequently return to their low-latitude breeding and calving grounds in the winter. Based on catch records corrected for illegal Soviet whaling, some 200 000 humpback whales were killed by pelagic whaling operations around Antarctica after 1900

(Clapham and Baker 2002, Allison 2010), driving a massive population decline (from an estimated pre-whaling population of 125 000) of this species.

The impact of whaling and the recovery of whale populations is a key focus of the IWC scientific committee. This committee takes historical population trends and combines them with population dynamic models to predict recovery rates. These models require historical catch records and an ability to accurately allocate catch to a source breeding population, estimates of biological parameters such as population structure, and abundance estimates, all of which are subject to considerable uncertainty (Baker and Clapham 2004, Jackson *et al.* 2008).

For whales hunted in the southern ocean, where the mixing of two or more breeding populations is suspected, the application of population dynamic models is particularly problematic. First pass attempts to allocate catches to a source population have been made using individual catch data collated and coded by the IWC from commercial whaling operations in Antarctic waters (Allison 2010). For example, the first model (denoted 'Naïve') assumed that the breeding populations corresponded to a single feeding area along arbitrary lines of longitude (Areas I-VI) (Mackintosh 1942, 1965, IWC 1998). As new information has emerged on the mixing of populations on the feeding grounds (Franklin *et al.* 2008, Steel *et al.* 2008, Gales *et al.* 2009, Anderson *et al.* 2010, Steel *et al.* 2011), alternative catch allocation models have been developed to include areas of mixing where whales are expected to be equally drawn from adjacent populations (e.g.,IWC 2010).

To date, assessments have been completed for several humpback whale breeding populations thought to have both simple and complex relationships between feeding areas and breeding grounds e.g. those whales that winter in the south-western Atlantic (Zerbini *et al.* 2006) and the south-eastern Pacific around Columbia, Panama and Equador (Johnston *et al.* In Press) versus those that winter along the west and east coast of South Africa (IWC 2009, 2011, 2012). The humpback whales that breed off eastern Australia and around the low latitude island groups of the South Pacific defined as 'Oceania' however, present unique challenges. This is largely due to uncertainties about both the population structure on the breeding grounds and the mixing of these populations in Antarctic waters.

As a further complication, recovery for the eastern Australian and Oceania populations has been variable. While the humpback whales migrating along eastern Australia have shown a high rate of population increase (10-11% per annum) (Noad *et al.* 2011), Oceania humpback whales are yet to show signs of recovery (Childerhouse and Gibbs 2006, Gibbs *et al.* 2006, Paton *et al.* 2006). This lack of recovery has prompted the relisting of the population as Endangered on the IUCN Redlist (IWC 1998, Childerhouse *et al.* 2008). The IWC Scientific Committee have therefore recommended an assessment of the mixing between eastern Australia and Oceania in Antarctic feeding Area V (130 - 180^{o}E - south of New Zealand and eastern Australia), as well as eastern Australia and western Australia in Antarctic feeding Area IV (80 – 130^{o}E - south of western Australia) (IWC 2011).

In light of the uncertainty about the population structure of humpback whales of Australia and the South Pacific, many different structure hypotheses have been proposed to simplify the comprehensive assessment (IWC 2006, 2011). . For the present study, we chose two different hypotheses that deal with this uncertainty by keeping the low latitude island groups of Oceania either separated or combined. Hypothesis 1 proposes a population structure where Oceania is sub-divided into four populations including New Caledonia, Tonga, Cook Islands and French Polynesia, with eastern Australia also considered a demographically independent population. This hypothesis was ranked ‘medium’ in plausibility as a realistic biological model during the workshop on the Comprehensive Assessment of Southern Hemisphere humpback whales, based on available biological evidence (IWC 2006, Olavarría *et al.* 2007, Garrigue *et al.* 2011). Hypothesis 2 combines New Caledonia, Tonga and French Polynesia to represent the ‘combined’ Oceania population, while retaining eastern Australia as a separate population and eliminating the Cook Islands. This hypothesis was proposed for priority consideration in the comprehensive assessment as a simple but plausible population structure scenario to deal with the limitations of catch allocation, which excludes the Cook Islands as it lacks a viable abundance estimate (Jackson *et al.* 2006, Jackson *et al.* 2009, IWC 2011). Although these hypotheses are largely arbitrary groupings and not necessarily representative of our best understanding of the true biology, they are presently an approximation that can be accommodated by the modelling process to develop catch allocation scenarios (IWC 2011).

Population genetic analysis has the potential to test these two hypotheses. Furthermore, genetic analysis can also assist in determining the degree of mixing of Australian and Oceania humpback whale populations on their feeding grounds. Both individual-level and population-level genetic methods hold promise for these tasks, i.e. using genetic information to ascertain population membership of individuals versus groups of individuals (Manel *et al.* 2005). Individual assignment tests use allele frequencies to assign individuals of unknown origin to their most likely source population and provide an estimate of statistical probability for each assignment (Davies *et al.* 1999). At the population level, mixed-stock analysis (MSA) uses allele frequencies in all potential contributing (baseline) populations and maximum likelihood or Bayesian methods to estimate proportional contributions of each population to a mixture (Pella and Milner 1987). Here, the question of interest is not the population origin of individual whales in a feeding ground mixture, but rather the population composition of a feeding ground mixture and how it changes in space and time. Two studies have employed MSA and maternally inherited mitochondrial DNA (mtDNA) haplotypes to allocate populations to Antarctic feeding areas but were missing data from Area V and eastern Australia respectively (Albertson-Gibb *et al.* 2008, Pastene *et al.* 2011). No study to date has combined mtDNA and nuclear markers to investigate humpback whale population allocation on the feeding grounds nor assessed statistical power to conduct a mixed-stock analysis.

In this study we draw on the most comprehensive dataset of mtDNA and nuclear microsatellite markers presently available for humpback whales of Australia and Oceania. The dataset stems from a large-scale collaborative effort between two laboratories, with a total of more than 1300 samples obtained over eleven years (Fig. 1). This extensive dataset provides us with an unprecedented opportunity to investigate the mixing of humpback whales on the feeding grounds, allowing us to fill a critical gap in knowledge.

Our specific objectives were to: 1) Assess the patterns and extent of genetic differentiation among the populations. 2) Apply a series of simulations to evaluate the power of the microsatellite and mtDNA datasets for individual assignment and mixed-stock analysis given available sample size and the patterns of genetic divergence among populations under the two hypotheses on population structure. 3) Extend the simulations to determine ways in which we can improve the accuracy and precision of mixed-stock analyses for these priority populations for future studies. 4) Estimate the population composition of Antarctic Area V samples collected during the Australia/New Zealand Antarctic Whale Expedition (AWE) in 2010, and interpret the findings in light of the simulation outcomes.

METHODS

The sample collection

I. Antarctic Area V sampling

Skin biopsy samples from 64 animals were collected from humpback whales in Antarctic feeding Area V (130 – 180^{o}E), south of New Zealand and eastern Australia (where the mixing of breeding populations is expected) during a six week Australian-New Zealand Antarctic Whale Expedition (AWE) conducted in February and March 2010. The majority of samples were collected from adult whales between 162˚E and 179˚E around the Balleny Islands. Samples were obtained using a biopsy dart propelled by a modified .22 calibre rifle and stored in 70% ethanol at -80˚C. All pods were sampled opportunistically with every effort made to sample all individuals within a pod, weather and time permitting.

Seven additional samples collected during IDCR/SOWER surveys of Antarctic Area V from 1999-2004 was also included in the sample dataset (Table 1) (described in Albertson-Gibb *et al.* 2008, Steel *et al.* 2008).

II. Source data

Mixed-stock analysis and assignment methods assume that all potential populations contributing to a mixture have been sampled. For the source dataset, we included six breeding/migratory populations which are likely to mix in Antarctic feeding Area IV or V both frequently and sporadically (Chittleborough 1965, Albertson-Gibb *et al.* 2008, Steel *et al.* 2008, Gales *et al.* 2009, Steel *et al.* 2011): western Australia, eastern Australia and Oceania (New Caledonia, Tonga, Cook Islands and French Polynesia). Oceania samples were obtained by members of the South Pacific Whale Research Consortium during synoptic surveys dating back to 1999 (described in Steel *et al.* 2008, Constantine *et al.* 2012). Australian samples were collected off Exmouth (Western Australia), Tasmania and Eden (NSW) between 2006 and 2008 (described in Schmitt *et al.* In Press). An additional 59 samples from eastern Australia were collected off Evan's Head (NSW), 560 kms north of Eden and were included in the second mixed stock analysis of the Area V samples (see METHODS section III).

See Table 1 for a summary of the sample details. Figure 1 provides a map of the sampling locations and the feeding areas.

Molecular Genetic Analysis

The DNA extraction, sex-typing, microsatellite genotyping and mtDNA sequencing have been described fully elsewhere. See Schmitt *et al*. (In Press) for the Australian samples, and Steel et al. (2008) and Constantine *et al*. (2012) for the samples from Oceania. Note that due to minor differences in the laboratory procedures, it was necessary to standardize the allele sizes for the microsatellite loci before the data sets could be combined. The standardization was achieved by re-analysis of 22 reference samples drawn from the Steel *et al*. (2008) study for which DNA was re-extracted and genotyped by the methods of Schmitt *et al.* (In Press). The lack of unresolvable discrepancies between the datasets allowed us to combine them for all subsequent analyses. The final combined microsatellite data set consisted of ten loci genotyped for 335 samples from Australia (plus 59 samples from Evans Head), 903 from Oceania and 62 from Antarctic Area V (Table 1).

For the mtDNA analysis, DNA sequences from the control region were truncated and aligned with a 470bp consensus region starting at position six of the reference humpback whale control region sequence (GenBank X72202: see Baker and Medrano-Gonzalez 2002, Olavarría *et al.* 2007). In total 312 sequences from the Australian samples, 872 from Oceania and 62 from Antarctic Area V were used in all subsequent analyses (Table 1), with 57 sequences from the Evan's Head samples used in the MSA estimation.

Statistical analysis

I. Genetic structure

For an initial evaluation of the two hypotheses for population structure; hypothesis 1 (H1) where the Oceania populations are considered separately, and hypothesis 2 (H2) where New Caledonia, Tonga and French Polynesia are combined (IWC 2006, Jackson *et al.* 2006, IWC 2011), we calculated genetic differentiation among each population pair for both scenarios using an Analysis of Molecular Variance (AMOVA Excoffier *et al.* 1992) as implemented in GenAlEx 6.5 (Peakall and Smouse 2006, 2012) with statistical testing by random permutation (999 permutations). For microsatellite data, an estimate of F_{ST} (infinite allele model) was calculated as per Weir & Cockerham (1984), Peakall *et al*. (1995) and Michalakis & Excoffier (1996). Given the high variability of microsatellite markers, Jost's D_{EST} (Jost 2008, Meirmans and Hedrick 2011), an unbiased estimator of divergence, was also calculated using a modified version of the R package DEMEtics V0.8.0 (Jueterbock *et al.* 2010), with statistical testing by bootstrapping with 1000 permutations. Compared with F_{ST}, D_{EST} partitions diversity based on the effective number of alleles rather than on the expected diversity to give an unbiased estimation of divergence (Jost 2008). For mtDNA data, an AMOVA was performed at both the nucleotide and haplotype level. For these analyses, genetic distance matrices were constructed using individual pairwise differences at all polymorphic nucleotide sites (following Excoffier et al. 1992), or haplotype differences among all individuals (Nei 1987). In keeping with the common practice in similar studies we use the notation F_{ST} for haplotype differentiation and Φ_{ST} for nucleotide differentiation (Olavarria *et al.* 2006, Olavarría *et al.* 2007, Rosenbaum *et al.* 2009).

II. Simulations

This is the first study to attempt a mixed stock analysis (MSA) of Antarctic area V humpback whale samples using both mtDNA and microsatellite markers. Even though we have assembled the most comprehensive genetic data set presently available for putative source populations, the sampling of the potential source population is not uniform across the study system. Therefore we employed computer simulations to assess the degree of confidence that we can have in our ability to assign an individual back to a source population and to estimate mixing on the feeding grounds.

In our computer simulations, the source data is analogous to a predictive model. A high degree of correct individual assignment or apportionment (see below for thresholds) would suggest that the level of genetic differentiation is sufficient to distinguish individuals from different populations or the populations themselves. The term 'confidence' comprises two components: accuracy and precision. We define 'accuracy' in these simulations as the agreement between the simulated value and the expected value, and 'precision' refers to the repeatability of the simulated value as measured by a confidence interval (i.e. the smaller the confidence interval, the more precise the value).

Three factors are known to be important for effectively estimating mixture proportions using MSA: the degree of differentiation among source populations and stocks, adequate sampling of all contributing source populations and a sufficient number of genetic markers (Pella and Milner 1987, Epifanio *et al.* 1995, Kalinowski 2004). Below we consider these factors.

Evaluating the resolution of source datasets for individual assignment and MSA

i. Individual assignment

To test our confidence in the source data to assign an individual whale to a given population we used the maximum likelihood classification in the program MLE (Topchy *et al.* 2004). This program can use both Mendelian (i.e. microsatellite) and non-Mendelian (i.e. mtDNA) loci simultaneously to assign individuals to the most likely source population based on the posterior probability for each classification. For mtDNA, the program uses haplotype frequencies at a single locus to obtain maximum likelihood estimates (Campbell *et al.* 2003). We applied this approach for three population mixtures likely to occur on the feeding grounds (i.e. western Australia and eastern Australia, eastern Australia and Oceania, and eastern Australia, New Caledonia, Tonga, Cook Islands and French Polynesia) using three genetic datasets: microsatellites, mtDNA, and microsatellites + mtDNA combined, employing a posterior probability assignment threshold of 0.90.

ii. Mixed-stock analysis (MSA)

For the MSA analysis we chose to use the program SPAM, v. 3.7 (Debevec *et al.* 2000, Alaska Department of Fish and Game. 2003), as it is currently the only mixed-stock simulation software that can accommodate both microsatellite and mtDNA data combined. SPAM employs a maximum likelihood approach, with allele frequency distributions modelled using the Rannala-Mountain posterior (Rannala and Mountain 1997). This approach allows for the estimation of allele frequencies for loci with many low-frequency alleles that can cause bias and/or imprecision in stock-composition estimates.

We used the 100% simulation feature in SPAM to assess the ability of MSA to accurately reapportion populations, assuming there is no mixing. This approach simulates a sample composed of 100% of each population from the source data and then attempts to reapportion simulated individuals to their population of origin. The simulation uses bootstrap resampling of allele frequencies from the source data based on a user nominated sample size of the 'pure stock'. This provides an initial benchmark with which to test the statistical power of our datasets for MSA and is a widely reported method in demonstrating the apportion accuracy for genetic stock identification applications (e.g. Smith *et al.* 2005, Beacham *et al.* 2006).

The simulations were performed on three separate datasets: microsatellites, mtDNA, and microsatellites + mtDNA combined. All 100% sample reapportioning simulations were conducted with 40 simulated individuals per iteration and 1000 bootstrap resamplings were used to calculate all mean proportional contribution estimates with 95% symmetric bootstrap confidence intervals. The sample size of 40 was chosen as a conservative sample number that might be collected on a six week voyage in the Southern Ocean. A population was considered identifiable if 90% or more of the simulated 'pure stock' was correctly identified to have originated from within the population (e.g. as demonstrated by Smith *et al.* 2005, Albertson-Gibb *et al.* 2008, Anderson *et al.* 2008, VanDeHey *et al.* 2010, Hess *et al.* 2011).

Improving the resolution of our source datasets for MSA

From (i) it was identified that combining microsatellites + mtDNA with Oceania populations grouped (H2) offered the most statistical power in a mixed-stock analysis and we therefore used this combination in subsequent MSA analyses. We examined the influence of source data sample size on the ability of MSA to accurately reapportion a 100% sample of each putative population by changing the number of samples in our source dataset to $N \geq 200$, 250, 300 and 400 for all populations (e.g. any population with less than 200 individuals in the source dataset would be increased to $N = 200$ and any population with $N \geq 200$ would remain at the current sample size) and repeated the 100% simulation in SPAM using 40 simulated individuals and 1000 bootstrap resamplings. When simulating an increase in source population sample sizes we used the original baseline allele frequencies to generate new individuals and therefore did not take into account the increased likelihood of rare alleles appearing if larger sample sizes were available. Despite the bias associated with sampling error (Anderson *et al.* 2008), the result will help determine whether the differences in sample size between populations have a strong influence on our ability to identify each population in a mixture. Larger baseline samples are expected to increase accuracy in population identification (Kalinowski 2004, Beacham *et al.* 2011).

Given extensive sampling in the southern ocean can be both expensive and time consuming, it is also useful to determine whether feeding ground sample size is important for an accurate MSA. We simulated different sample sizes of a 'pure stock', representing samples on the feeding grounds to determine the minimum number of samples required to confidently reapportion one population from its neighbouring population, likely to occur in a feeding ground mixture (i.e. western Australia and eastern Australia; Oceania and eastern Australia). Using the 100% simulation feature in SPAM we simulated various feeding ground sample sizes (i.e. N = 40, 65, 80, 120, 160, 200, 280, 400) and compared estimates of mean correct assignment of all individuals to their population of origin. By plotting these estimates for one population from each potential mixture and identifying the inflection point where the mean correct assignment begins to stabilize, we could determine the minimum sample size required for that given level of accuracy.

Realistically, humpback whale breeding populations on the Antarctic feeding grounds are likely to be a mixture of two neighbouring populations (e.g. 50% from western Australia and 50% from eastern Australia; 60% from eastern Australia and 40% from Oceania). By simulating realistic population mixtures with user defined proportions in SPAM, we can assess the degree of confidence in the source dataset to determine the proportion of humpback whales assigned to each population in a mixture. Predicted precision and accuracy was assessed from the difference between predicted and observed population proportions. We used an error rate of ≤ 10% outside the predicted proportion to imply confidence in the source data in predicting stock contribution in a mixture (VanDeHey *et al.* 2010). We simulated mixtures of N = 40 for both marginal feeding areas with expected proportional contributions varying from 0.0 to 1.0. We also investigated the influence of the sample size of our source data and the number of simulated individuals (mixture or feeding ground sample size) on our ability to estimate different population contributions in the two mixtures by repeating the analysis using a population sample size of $N \geq 200$ and a mixture sample size of N = 160 based on results from our previous analyses (ii and iv).

III. Mixed-stock estimation of the 62 individual samples collected from Antarctic feeding Area V

We used the program SPAM and our source data to calculate maximum likelihood estimates of humpback whale population contributions (with Oceania populations grouped and not grouped) to our Antarctic feeding Area V samples using all three genetic marker sets and 10000 bootstrap resamplings. Based on the outcome of the mixed-stock analysis simulations (see Results II ii), we conducted the estimation both with and without the 59 samples from Evan's Head in eastern Australia. We chose not to conduct an individual assignment as most individual whales from the source data could not be assigned to a population of origin (see Results III).

RESULTS

In total we have assembled a data set consisting of 1309 individuals, with data for the microsatellite loci available for 1300 samples, while mtDNA sequences were available for 1246 samples. The sex ratio was significantly biased towards males (743 males to

514 females, χ^2 = 54.6, P < 0.0001) for all sampling locations with the exception of the Cook Islands (47 males to 42 females, χ^2 = 0.3, P = 0.60) (Table 1).

Statistical Analysis

Summary genetic data for each microsatellite locus as well as variability in the mtDNA control region for all samples are presented in Appendix 1 and 2. None of the ten microsatellite loci violated Hardy-Weinberg equilibrium assumptions after sequential Bonferroni correction (Appendix 1). There was some evidence for a non-random association of alleles between EV14 and rw4-10, and EV37and GT23 for New Caledonia but not elsewhere.

I. Genetic structure

In evaluating the evidence in support of the two alternative structure hypotheses we considered the patterns of genetic diversity for the Oceania population sub-divided (H1) versus grouped (H2).

Across the whole data set with Oceania sub-divided (H1), genetic differentiation was weak but significant, (microsatellites: F_{ST} = 0.004, P = 0.001; D_{EST} = 0.004, P = 0.035, mtDNA: haplotype level F_{ST} = 0.018, P = 0.001; nucleotide level Φ_{ST} = 0.025, P = 0.001). Pairwise comparisons for the ten microsatellite loci detected no significant difference between Antarctic Area V and eastern Australia (infinite allele model of mutation F_{ST} and Jost's D_{EST} = 0, P > 0.05), and Antarctic Area V and New Caledonia (F_{ST} and D_{EST} < 0.005, P > 0.05). Using the D_{EST} index there was also no significant structure detected between the Cook Islands and New Caledonia, Tonga, and French Polynesia respectively (P > 0.05) (Table 2a). For mtDNA there was low but significant structure detected at the haplotype level between all comparisons except Antarctic Area V and eastern Australia, and French Polynesia and the Cook Islands (P > 0.05). At the nucleotide level twelve pairwise comparisons were significantly different (P < 0.05) and nine were not including Antarctic Area V and eastern Australia, New Caledonia, Tonga and the Cook Islands respectively (P > 0.05) (Table 3a).

With the Oceania populations grouped (H2), genetic differentiation was similarly weak but significant across the entire dataset, but with a substantial increase in D_{EST} (microsatellites: D_{EST} = 0.017, P = 0.001, mtDNA: haplotype level F_{ST} = 0.012, P = 0.001; nucleotide level: Φ_{ST} = 0.020, P = 0.001). Pairwise genetic differentiation was weak but significant between all comparsions (P = 0.001) except Antarctic Area V and eastern Australia for all statistics, and Antarctic Area V and Oceania for Φ_{ST} (Table 2b and 3b).

II. *Simulations*

Evaluating the resolution of source datasets for individual assignment and MSA

i. *Individual assignment*

At the individual level, we could re-assign more whales correctly to a population of origin using the combined microsatellite + mtDNA dataset than microsatellites or mtDNA alone (Table 4). However, a large proportion of individuals remained unassigned using the posterior probability threshold of > 0.90 .

For the western Australia/eastern Australia mix, 45.9% of all individuals could be re-assigned to a population correctly for the combined marker set compared to 30.7% for microsatellites and 40.3% for mtDNA alone. More whales from western Australia could be re-assigned for all marker sets than individuals from eastern Australia.

We then evaluated the outcomes of individual assignment tests separately for the two alternative population structure hypotheses for Oceania. When populations of Oceania were grouped (H2) for the eastern Australia/Oceania mix, 43.3% of individuals could be re-assigned correctly using microsatellites + mtDNA, with only 25.7% assigned correctly for microsatellites and 22.9% for mtDNA alone. More individuals from Oceania were correctly re-assigned for all marker sets.

When populations of Oceania were retained as separate groups (H1) significantly fewer individuals could be re-assigned correctly when compared with the groupings under H2, [ranging from only 1.0% for microsatellites to 5.8% for the combined marker set ($F_{1,5}$ = 18.0, P = 0.013)]. More individuals from eastern Australia could be re-assigned correctly for all marker sets than the other populations with an inability to re-assign any individuals to the Cook Islands.

ii. Mixed Stock Analysis (MSA)

Overall, the mtDNA alone reapportions a 'pure stock' with a slightly greater accuracy than did the combined and microsatellite datasets (Fig. 2). The combined mtDNA + microsatellite dataset however, showed a consistent trend of more precise estimates (characterised by smaller Coefficient of Variation, *CV*) (e.g. for populations grouped under H2, mtDNA mean *CV* = 0.06; mtDNA + microsatellites mean *CV* = 0.05). Overall, confidence in reapportioning a 'pure stock' was highest when populations were grouped as for H2. Under this grouping mixed stock apportionment was correct 90% or more of the time for mtDNA and the combined marker set with the exception of eastern

Australia (average ~ 94%). By contrast when populations were grouped as for H1, correct apportionment was lower than the 90% threshold (between 77.2% to 86.7%, with the exception of western Australia, eastern Australia and French Polynesia for mtDNA, and western Australia and New Caledonia for the combined marker set). Given the improved confidence under the H2 groupings and consistent trend of improved precision (lower confidence intervals), we conducted all subsequent simulations using the combined marker set and grouped as for H2 (with Oceania populations combined).

Improving the resolution of our source datasets for MSA

The cost involved in obtaining whale samples is very high. Simulations allowed us to evaluate the potential benefits of increasing sample size from putative breeding populations and on the feeding grounds under the assumptions that present allele frequencies are representative of the source populations.

Increasing the simulated sample size for eastern Australia to 200 had the greatest effect on our ability to reapportion a ‘pure stock’ using MSA, increasing the mean correct assignment in the 100% sample reapportioning simulations from 88.5% to 93.9% (an increase of 5.4%) (Fig. 3). Accuracy and precision of assignment improved only marginally for western Australia and eastern Australia when sample size was increased to ≥ 250, 300 and 400 respectively (WA by 1.7% and EA by 3.1% overall).

The number of simulated individuals, representing feeding ground individuals, had minimal impact on the accuracy of reapportioning a ‘pure stock’ using MSA, although precision increased (Fig. 4a and b). Confidence intervals around the mean correct assignment begin to stabilize at a sample size of 160 for both feeding area mixes.

In simulating mixtures of different proportions using our original sample sizes and a mixture sample size of *N* = 40, the differences between expected and estimated proportions for both feeding ground mixes were all within the 10% threshold error rate (3.4 to 3.6%) for identity except when eastern Australia was at an expected proportion of 90% (11.3%) (Fig. 5a and b). Confidence intervals however, were all outside the 10% threshold error rate and at their largest for a 50:50 mixture.

When sample size was increased from 131 to 200 for eastern Australia the accuracy of proportion estimates increased for both mixtures. Differences between the expected and estimated proportions of western Australia (and likewise for eastern Australia) were reduced by an average of 0.9% (all differences within 1.8%) while for eastern Australia (and likewise Oceania), the differences were reduced by an average of 3.0% (all differences within 6.3%) with estimates improving by as much as 5% for expected proportions of 50-90% (Fig. 5c and d). Increasing the source data sample size had little effect on confidence intervals.

When mixture sample size was also increased to *N* = 160, confidence intervals were reduced by half and were within the 10% threshold error rate for all expected proportions for both feeding area mixtures (Fig. 5e and f).

III. Mixed-stock estimation of the 62 individual samples collected from Antarctic feeding Area V

With populations grouped as for H2 eastern Australia accounted for 46.8% (SE = 5.9%), Oceania accounted for 52.4% (SE = 6.7%) and western Australia, 0.8% (SE = 0.1%) of the 62 Antarctic feeding Area V samples using the combined mtDNA + microsatellite dataset. Eastern Australia accounted for as much as 68.9% (SE = 10.8%)

using the mtDNA only dataset (Oceania: 31.1% ,SE = 10.8%) but only 29.6% (SE = 3.7%) using the microsatellite dataset (Oceania: 70.4%, SE = 8.9%) with no contribution from western Australia (Fig. 6a).

Although signals in the above simulations appear to favour a grouping of populations in Oceania, under H1, three populations for mtDNA and two for the combined dataset performed above the 90% threshold in the 100% sample reapportioning simulations. Based on this result we decided to also perform an MSA with Oceania as separate populations. In this case, contributions from Oceania to the Antarctic samples were found to be predominantly drawn from New Caledonia across all three marker sets (combined: 50.0%, SE = 6.4%; mtDNA: 27.9%, SE = 10.6%; microsatellites: 68.7%, SE = 8.6%) with close to negligible contributions from Tonga (Fig. 6b).

At the beginning of the simulation study, the 59 samples from Evan’s Head (eastern Australia) were not available, and were therefore not included in the simulations. As increasing the number of samples from eastern Australia was found to boost the accuracy of population allocation estimates, we repeated the MSA for the Area V samples with these additional 59 samples. An AMOVA analysis between the three eastern Australian sampling locations found no significant differentiation for either the microsatellite (F_{ST} = 0.000, P = 0.5; D_{EST} = 0.000, P = 0.5) or the mtDNA (F_{ST} = 0.003, P = 0.2; Φ_{ST} = 0.000, P = 0.5) datasets, thereby justifying the pooling of the data. The addition of these 59 samples had little effect on apportionment for the Area V samples using the mtDNA data only however, the estimated contribution of eastern Australia for both microsatellite datasets increased, while the contribution of New Caledonia and Oceania as a whole decreased (combined: EA, 53.2%, SE = 6.8%; NC, 43.7%, SE =

5.5%; Oceania, 46.8%, SE = 5.9%; microsatellites: EA, 40.6%, SE = 5.1%; NC = 49.6%, SE = 6.3%; Oceania = 56.1%, SE = 7.1%) (Fig. 6c and d).

DISCUSSION

In this study we combined genetic analysis of mtDNA and microsatellite DNA data and simulation to explore the role of differentiation among source populations, as well as source and feeding ground sample size in effectively identifying pure samples and estimating mixture proportions. Consistent with other studies weak but significant differentiation was detected between populations. In cases such as this where differentiation is low, simulations are particularly important to evaluate the statistical power given the available genetic markers and the sample size. In the discussion that follows we consider the strengths and limitations of our current datasets for drawing conclusions about population allocations on the feeding grounds and offer recommendations on sampling gaps and adjustments that can be implemented to ensure our estimates are robust.

Genetic structure

This is the first study to assess the patterns of genetic differentiation at nuclear loci across the populations of Australia and Oceania. The degree of differentiation among pairwise comparisons was consistently low for both marker sets (microsatellites: H1 D_{EST} from 0.000 to 0.035, H2 D_{EST} from 0.000 to 0.031; mtDNA: H1 F_{ST} from 0.002 to 0.037, H2 F_{ST} from 0.002 to 0.016; Table 2 and 3). Differentiation was particularly weak among the ungrouped populations of Oceania (H1), with a D_{EST} of only 0.009 between New Caledonia and Tonga, 0.000 between Tonga, the Cook Islands and French

Polynesia and 0.009 between Tonga and French Polynesia (mtDNA F_{ST} = 0.009, 0.010, 0.003 and 0.017 respectively). In spite of this result, differentiation between New Caledonia, Tonga and French Polynesia was found to be significant for both markers across all statistics (P < 0.05), suggesting demographic independence among these breeding grounds. These results are consistent with recent findings based on multi-state measurements of genotype exchange within the ungrouped populations of Oceania and between these populations and eastern Australia i.e. eastern Australia may not be more isolated from Oceania than animals within Oceania are from one another (Jackson *et al.* 2012).

Individual assignment tests

Using individual assignment tests, we could assign more whales to a population using the combined microsatellite + mtDNA dataset than microsatellites or mtDNA alone. However, a large proportion of individuals remained unassigned (50 - 52% under H2 grouping and 91% under H1) using the posterior probability threshold of > 0.90. The poor performance of the individual assignment tests is likely to be a combination of low differentiation and an inadequate number of genetic markers and samples. The ability to assign individuals to the correct population of origin is known to break down when the populations exhibit low differentiation and there is no *a priori* population information (F_{ST} < 0.03) (Latch *et al.* 2006). With the addition of substantially more loci, we may improve our ability to distinguish between populations characterised by low differentiation and ultimately, our success in individual assignment (Allendorf *et al.* 2010). The populations of Oceania (New Caledonia, Tonga, Cook Islands and French Polynesia) were the least differentiated in the AMOVA analysis (see last section) and also had the least number of individuals correctly assigned to them. Significantly more individuals could be correctly assigned however, when Oceania populations were combined, creating a much larger sample size. Similarly, western Australia was the most distinguishable population in both the individual assignment tests and the AMOVA, with the highest level of differentiation when compared to a neighbouring stock (WA vs EA: mtDNA F_{ST} = 0.042; microsatellites D_{EST} = 0.031).

Feasibility of a mixed stock analysis

Our ability to confidently reapportion the populations of Australia and Oceania using a mixed-stock analysis was influenced by the choice of genetic markers and the degree of differentiation. Our 100% sample reapportioning simulation results suggest that the ten microsatellite loci alone did not allow us to accurately reapportion each population using a 90% identity threshold, and may therefore give a poor estimate of population apportionment on the feeding grounds. This finding is likely to have been influenced by the very low differentiation at the microsatellite loci. Although the mtDNA control region dataset alone could accurately discriminate among most populations at the 90% identity threshold and a mixture sample size of 40, the combined mtDNA + microsatellite dataset could do so with a greater precision (smaller confidence intervals). Given the increase in precision was not significant however, mtDNA control region sequences alone may be sufficient for MSA in humpback whales. These results are consistent with the highly discriminatory nature of mtDNA due to their haploid nature and uniparental inheritance which are expected to result in a larger genetic drift compared to nuclear loci (e.g. Avise 1995, Sunnucks 2000). Although nuclear loci can provide greater resolution in discriminating between populations due to the variable nature of some markers (Angers and Bernatchez 1998, Selkoe and Toonen 2006), the AMOVA results in this study suggest that the ten microsatellite loci may add little to the discriminatory power of mtDNA. Therefore, given the low differentiation that characterizes Oceania and Australian humpback whales at the nuclear level, it may unnecessary to combine both marker types to estimate the mixture contributions of populations on the feeding grounds.

100% sample reapportioning simulations showed that western Australia, eastern Australia, New Caledonia and French Polynesia could be reapportioned above the 90% threshold for either the mtDNA or combined dataset using original sample sizes and a simulation sample size of 40. The lack of power to reapportion Tonga and the Cook Islands is likely to be a consequence of the interplay between the smaller sample size of the Cook Islands and the weak differentiation between them. Yet when the Cook Islands was removed from the simulation analysis, the ability to reapportion Tonga increased from 86% to 89% and French Polynesia increased from 89% to 92% using the combined mtDNA + microsatellite dataset (data not shown). This result and the lack of significant differentiation between Tonga, the Cook Islands and French Polynesia, is consistent with the suggestion that the Cook Islands aggregation is transient based on strong connections with Tonga (Garrigue *et al.* 2002, Hauser *et al.* 2010). Our results confirm that the region should not be included as a unique entity in comprehensive assessments, as has been suggested previously (IWC 2006).

Performance of mixed-stock simulations under the two hypotheses for population structure

Based on the results of the 100% sample reapportioning simulations and despite the low differentiation between New Caledonia, Tonga and French Polynesia, with the exclusion of the Cook Islands, we suggest accurate estimates of humpback whale feeding ground composition will be possible under both hypothesis 1 and 2 given adequate sample sizes.

Improving the accuracy and precision of mixed-stock analyses

Sample size of the source populations was found to be a key factor influencing the accuracy of population identification. For example, when sample sizes were increased through simulation, greater confidence in reapportionment to eastern Australia was observed with a marginal improvement for western Australia. A moderate increase in the sample size of eastern Australia from 131 to 200 had the greatest effect on our ability to reapportion the population, increasing the accuracy from 88.5% to 93.9%; above the 90% threshold for identity. Increasing source sample sizes beyond 200 produced diminishing returns. Fisheries stock simulation studies generally support these findings, reporting little benefit in increasing population sample size beyond 100 or 200 fish, a result which is dependent on the resolution of the genetic markers set used (Kalinowski 2004, VanDeHey *et al.* 2010, Hess *et al.* 2011). Indeed a moderate increase in source sample size has been found to have greater gains in the statistical power of the data than moderate increases in the number of nuclear loci (up to 20 loci; 8-33 alleles each), particularly when F_{ST} is low (< 0.01) and the average correct assignment to a populations is greater than 80% (Kalinowski 2005, Morin *et al.* 2009, Hess *et al.* 2011).

The simulations demonstrate that increasing the sample size of eastern Australia to 200 and the feeding grounds to 160 produced estimates and confidence intervals within the stringency threshold for all expected proportions for both feeding area mixtures. Therefore a moderate increase in both source and feeding ground sample sizes is recommended from this study for accurate apportionment of these populations on the Antarctic feeding grounds and an entirely viable option given prior sampling efforts (Pastene *et al.* 2011, Steel *et al.* 2011).

Simulation limitations

There are two ways in which our analyses may bias estimates of simulation performance for all three marker sets. Firstly, SPAM has been found to overestimate the predicted accuracy and precision of mixed-stock analysis by resampling from the baseline with replacement, particularly for closely related populations (Anderson *et al.* 2008). Different types of simulation software have attempted to address this problem (Banks and Eichert 2000, Piry *et al.* 2004, Anderson *et al.* 2008) but there is still no consensus on the most robust way to demonstrate the statistical power of a baseline using simulations. There is also no software yielding unbiased estimates of genetic stock identification that can accommodate combined mtDNA and nuclear data. Preliminary analyses using the 100% sample reapportioning simulation feature in ONCOR which attempts to reduce this bias (Anderson *et al.* 2008) and the microsatellite data only, produced similar estimates when mean correct assignment was greater than 90% (e.g. SPAM estimate for Oceania under H2 = 94.1%; ONCOR estimate = 92.4%). Nonetheless, it is uncertain what effect sampling the baseline with replacement might have on mtDNA or combined mtDNA and nuclear data. Despite the bias associated with SPAM, our results provide information on the relative accuracy and precision of marker types, and the effect increasing marker number and sample size has on genetic stock identification in humpback whales. Secondly, it should be noted that all MSA simulations assume that the source samples and allele frequency estimates are representative of the populations present in the mixture and, therefore, do not take account of unrepresentative baseline samples and alleles or omitted source populations. As there is no way to systematically account for the possibility of individuals from unsampled populations in the mixture, simulation and estimation results should be interpreted with a degree of caution.

Population composition of Area V samples

Our simulations offer important clues about the degree of confidence we can expect in estimating the population apportionment of Australia and Oceania humpback whales on the Antarctic feeding grounds using a mixed-stock analysis. It is evident that these estimates are influenced by the genetic distinctiveness among source populations, choice of genetic markers, and both source and mixture sample sizes.

Notwithstanding the value of increasing sample sizes, the simulations indicate that our mtDNA and combined mtDNA + microsatellite datasets have the potential to provide estimates that are close to satisfying the 90% stringency threshold for both hypotheses of population structure.

In light of these simulations, what can we safely conclude about the population mixture of our Antarctic Area V sample? Two important insights emerge from the simulations: 1. The contribution of western Australia is small to negligible. We have a high degree of confidence in this conclusion given both the small MSA estimated error (Fig. 6a to d) and the outcomes of the simulations that predict close agreement between estimated and expected proportion mixtures when the contribution from western Australia is low (Fig. 5b). 2. Both eastern Australia and Oceania (largely represented by New Caledonia) make substantial contributions to the Area V sample. What is less certain is the exact apportionment of feeding Area V samples to these populations. This uncertainty arises from the discrepancy between the MSA estimates between the three different marker sets, and the error surrounding these estimates (Fig. 6a and b). The simulations also indicated the largest errors are predicted when there are more or less even contributions of eastern Australia and Oceania (Fig. 5a). With the addition of the 59 Evan's Head samples to eastern Australia however, the discrepancy between estimates for each dataset was reduced (Fig. 6c and d). Thus, while we can be confident that there is a substantial contribution of both populations to Antarctic feeding Area V, whether or not the contributions are even or a little biased will perhaps require more discriminating genetic markers.

Collectively, our results add support to individual connection studies using *Discovery* tags, photo identification and genotype matches, as well as satellite tags linking eastern Australia to Antarctic Area V (Dawbin 1966, Olavarria *et al.* 2006, Rock 2006, Franklin *et al.* 2008, Gales *et al.* 2009, Constantine *et al.* 2011, Steel *et al.* 2011). From the Antarctic Area V whales used in this study, Constantine *et al.* (2011) and Steel et al. (2011) found the majority of fluke and genotype matches were to eastern Australia, with one match to New Caledonia and the New Zealand migratory corridor. Despite the low number of matches with Oceania, their result nonetheless implies that Area V is a region where mixing between eastern Australia and Oceania occurs. Their study also corroborates our findings of a negligible contribution from western Australia in Area V, with no matches found. This is consistent with the results of *Discovery* tag recoveries, implying humpback whales breeding in western Australia are more likely to mix with those of eastern Australia east of 115°E in Antarctic Area IV (Chittleborough 1965, Dawbin 1966). Our findings also support the results of other recent MSA studies, that despite their data limitations offered new evidence that New Caledonia may have a larger connection with Area V than previously thought (Albertson-Gibb *et al.* 2008, Pastene *et al.* 2011).

This study is the first to comprehensively assess the power of a combined microsatellite and mtDNA genetic dataset to discriminate between Australian and Oceania humpback whales on the Antarctic feeding grounds at both individual and population levels. We emphasize that these results should be considered preliminary, pending a re-analysis of other population structure scenarios as new biological evidence comes to light. Despite the need for more samples to improve estimates of population allocation, our study strengthens the emerging genetic and non-genetic evidence that Antarctic feeding Area V is utilized by humpback whales from both eastern Australia and the more vulnerable population of New Caledonia, representing Oceania.

ACKNOWLEDGEMENTS

Thanks to the Australian Marine Mammal Centre at the Australian Antarctic Division for their generous financial support of the project, the South Pacific Whale Research Consortium for providing genetic data from New Caledonia, Tonga, the Cook Islands and French Polynesia and Michele Masuda from Alaska Fisheries Science Centre, NOAA Fisheries, Alaska for her advice on mixed-stock analysis simulations. The Antarctic Whale Expedition (AWE) was generously funded by the Australian and New

Zealand Governments and had logistical support from the Australian Antarctic Division and National Institute for Water and Atmospheric research.

LITERATURE CITED

Alaska Department of Fish and Game. 2003. SPAM Version 3.7: Addendum II to user's guide for version 3.2, Division of Commercial Fisheries, Gene Conservation Laboratory, Special Publication No. 15, Anchorage, Alaska, USA.

Albertson-Gibb, R., C. Olavarria, C. Garrigue, N. Hauser, M. Poole, L. Florez-Gonzalez, C. Antolik, et al. 2008. Using mitochondrial DNA and mixed-stock analysis to estimate migratory allocation of humpback whales from Antarctic feeding areas to South Pacific breeding grounds. Paper SC/60/SH15 presented at the IWC Scientific Committee, Santiago, Chile.

Allendorf, F.W., P.A. Hohenlohe and G. Luikhart. 2010. Genomics and the future of conservation genetics. Genetics 11:697-709.

Allison, C. 2010. IWC individual catch database Version 5.0; Date: 1 October 2010.

Anderson, E.C., R.S. Waples and S.T. Kalinowski. 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences 65:1475-1486.

Anderson, M., D. Steel, T. Franklin, W. Franklin, D. Paton, D. Burns, P. Harrison, et al. 2010. Microsatellite genotype matches of eastern Australian humpback whales to Area V feeding and breeding grounds. Paper SC/62/SH7 presented at the IWC Scientific Committee, Agadir, Morocco.

Angers, B. and L. Bernatchez. 1998. Combined use of SMM and non-SMM methods to infer fine structure and evolutionary history of closely related brook charr (*Salvelinus fontinalis*, Salmonidae) populations from microsatellites. Molecular Biology & Evolution 15:143-159.

Avise, J.C. 1995. Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. Conservation Biology 9:686-690.

Baker, C.S. and P.J. Clapham. 2004. Modelling the past and future of whales and whaling. Trends in Ecology & Evolution 19:365-371.

Baker, C.S. and L. Medrano-Gonzalez. 2002. World-wide distribution and diversity of humpback whale mitochondrial DNA lineages. Pages 84-99 *in* C.J. Pfeiffer, ed. Molecular and Cell Biology of Marine Mammals. Krieger Publishing, Melbourne, FL.

Banks, M.A. and W. Eichert. 2000. WHICHRUN (version 3.2): A computer program for population assignment of individuals based on multilocus genotype data. Journal of Heredity 91:87-89.

Beacham, T.D., J.R. Candy, K.L. Jonsen, J. Supernault, M. Wetklo, L.T. Deng, K.M. Miller, et al. 2006. Estimation of stock composition and individual identification of Chinook salmon across the Pacific Rim by use of microsatellite variation. Transactions of the American Fisheries Society 135:861-888.

Beacham, T.D., B. McIntosh and C.G. Wallace. 2011. A comparison of polymorphism of genetic markers and population sample sizes required for mixed-stock analysis of sockeye salmon (Oncorhynchus nerka) in British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 68:550-562.

Beacham, T.D., M. Wetklo, C. Wallace, J.B. Olsen, B.G. Flannery, J.K. Wenburg, W.D. Templin, et al. 2008. The application of Microsatellites for stock identification of Yukon River Chinook salmon. North American Journal of Fisheries Management 28:283-295.

Cadrin, S.X., K.D. Friedland and J.R. Waldman. 2005. Stock identification methods: applications in fishery science. Elsevier, Inc, Burlington, MA.

Campbell, D., P. Duchesne and L. Bernatchez. 2003. AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. Molecular Ecology 12:1979-1991.

Childerhouse, S. and N. Gibbs. 2006. Preliminary report for the Cook Strait Humpback Whale Survey 2006. Unpublished report to the Department of Conservation, New Zealand.

Childerhouse, S., J. Jackson, C. Baker, N. Gales, P. Clapham and R.L. Brownell Jr. 2008. *Megaptera novaeangliae* (Oceania subpopulation) IUCN 2012. IUCN Red List of Threatened Species. Version 2012.1.

Chittleborough, R.G. 1965. Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). Australian Journal of Marine Freshwater Research 16:33-128.

Clapham, P.J. and C.S. Baker. 2002. Modern whaling. Pages 1328-1332 *in* W.F. Perrin, B. Wursig and J.G.M. Thewissen, eds. Encyclopedia of Marine Mammals. Academic Press, New York.

Constantine, R., J. Allen, P. Beeman, D. Burns, J. Charrassin, S. Childerhouse, M.C. Double, et al. 2011. Comprehensive photo-identification matching of Antarctic Area V humpback whales. Paper SC/63/SH16 presented at the IWC Scientific Committee, Tromso, Norway.

Constantine, R., J.A. Jackson, D. Steel, C.S. Baker, L. Brooks, D. Burns, P. Clapham, et al. 2012. Abundance of humpback whales in Oceania using photo-identification and microsatellite genotyping. Marine Ecology Progress Series 453:19.

Davies, N., F.X. Villablanca and G.K. Roderick. 1999. Determining the source of individuals: multilocus genotyping in non-equilibrium population genetics. Trends in Ecology & Evolution 14:17-21.

Dawbin, W.H. 1966. The seasonal migratory cycle of humpback whales. Pages 145-170 *in* K.Norris, ed. Whales, Dolphins and Porpoises. University of California Press, Berkeley, California.

Debevec, E.M., R.B. Gates, M. Masuda, J. Pella, J. Reynolds and L.W. Seeb. 2000. SPAM (Version 3.2): Statistics program for analyzing mixtures. Journal of Heredity 91:509-510.

Epifanio, J.M., P.E. Smouse, C.J. Kobak and B.L. Brown. 1995. Mitochondrial DNA divergence among populations of American shad (Alosa sapidissima): how much variation is enough for mixed-stock analysis? Canadian Journal of Fisheries and Aquatic Science 52:1688-1702.

Excoffier, L., P.E. Smouse and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Franklin, T., F. Smith, N. Gibbs, S. Childerhouse, D. Burns, D. Paton, W. Franklin, et al. 2008. Migratory movements of humpback whales (*Megaptera novaeangliae*) between eastern Australia and the Balleny Islands, Antarctica, confirmed by photo-identification. Report (SC/60/SH2) to the Scientific Committee of the International Whaling Commission, Santiago, Chile.

Gales, N., M.C. Double, S. Robinson, C. Jenner, M. Jenner, E. King, J. Gedamke, et al. 2009. Satellite tracking of southbound East Australian humpback whales (*Megaptera novaeangliae*): challenging the feast or famine model of migrating whales. Paper SC/61/SH17 presented at the IWC Scientific Committee, Madeira, Portugal.

Garrigue, C., A. Aguayo, V.L.U. Amante-Helweg, C.S. Baker, S. Caballero, P. Clapham, R. Constantine, et al. 2002. Movements of humpback whales in Oceania, South Pacific. Journal of Cetacean Research and Management 4:255-260.

Garrigue, C., R. Constantine, M. Poole, N. Hauser, P. Clapham, M. Donoghue, K. Russel, et al. 2011. Movement of individual humpback whales between the breeding grounds of Oceania, South Pacific 1999-2004. Journal of Cetacean Research and Management (Special Issue) 3.

Gibbs, N., D. Paton, S. Childerhouse and P. Clapham. 2006. Assessment of the current abundance of humpback whales in the Lomaiviti Island Group of Fiji and a comparison with historical data.

Hauser, N., A. Zerbini, I. Geyer, M.P. Heide-Jorgensen and P. Clapham. 2010. Movements of satellite-monitored humpback whales, *Megaptera novaeangliae*, from the Cook Islands. Marine Mammal Science 26:679-685.

Hess, J.E., A.P. Matala and S.R. Narum. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Molecular Ecology Resources 11 (Suppl. 1):137-149.

IWC. 1998. Annex G. Report of the Sub-Committee on the Comprehensive Assessment of Southern Hemisphere Humpback Whales 48:170-82. International Whaling Commission.

IWC. 2006. Report of the Workshop on the Comprehensive Assessment of Southern Hemisphere Humpback Whales. SC/58/Rep 5.

IWC. 2009. Annex H. Report of the sub-committee on other Southern Hemisphere whale stocks (SH). Journal of Cetacean Research and Management 11(SUPPL.):220-247.

IWC. 2010. Annex H. Report of the sub-committee on other Southern Hemisphere whale stocks. Journal of Cetacean Research and Management 11 (SUPPLE. 2):218-251.

IWC. 2011. Annex H. Report of the sub-committee on other Southern Hemisphere whale stocks. Journal of Cetacean Research and Management 12 (SUPPL.):203-226.

IWC. 2012. Annex H. Report of the sub-committee on other Southern Hemisphere whale stocks. Journal of Cetacean Research and Management 13 (SUPPL.):192-216.

Jackson, J., M. Anderson, D. Steel, L. Brooks, P. Baverstock, D. Burns, P. Clapham, et al. 2012. Multistate measurements of genotype interchange between East Australia and Oceania (IWC breeding sub-stocks E1, E2, E3 and F) between 1999 and 2004. Paper SC/64/SH22 presented at the IWC Scientific Committee, Panama City, Panama.

Jackson, J., A. Zerbini, P. Clapham, C. Garrigue, N. Hauser, M. Poole and C.S. Baker. 2006. A Bayesian assessment of humpback whales on breeding grounds of eastern Australia and Oceania (IWC Stocks, E1, E2, E3 and F). Paper SC/A06/HW52 presented to the International Whaling Commission Comprehensive Assessment of Southern Hemisphere humpback whales, 2006, Hobart, Australia.

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

Jackson, J.A., A. Zerbini, P. Clapham, R. Constantine, C. Garrigue, N. Hauser, M.M. Poole, et al. 2009. Update to SC/60/SH14: Progress on a two-stock catch allocation model for reconstructing population histories of east Australia and Oceania. Paper SC/F09/SH8 submitted to the IWC Intersessional Workshop on Southern Hemisphere Humpback Whale Modelling Assessment Methodology (unpublished).

Johnston, S.J., A.N. Zerbini and D.S. Butterworth. In Press. A Bayesian Approach to assess the status of Southern Hemisphere humpback whales (*Megaptera novaeangliae*) with an application to Breeding Stock G. Journal of Cetacean Research and Management (Special Issue).

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

Jueterbock, A., P. Kraemer and G.D.J. Gerlach. 2010. *DEMEtics v0.8.0*. University of Oldenburg, Oldenburg, Germany.

Kalinowski, S.T. 2004. Genetic polymorphism and mixed-stock fisheries analysis. Canadian Journal of Fisheries & Aquatic Sciences 61:1075-1082.

Kalinowski, S.T. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? Heredity 94:33-36.

Latch, E.K., G. Dharmarajan, J.C. Glaubitz and O.E. Rhodes Jr. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics 7:295-302.

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

Mackintosh, N.A. 1965. The stock of whales. Fishing News (Books) Ltd., London.

Manel, S., O.E. Gaggiotti and R.S. Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology and Evolution 20:136-142.

Meirmans, P.G. and P.W. Hedrick. 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11:5-18.

Michalakis, Y. and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061-1064.

Morin, P.A., K.K. Martien and B.L. Taylor. 2009. Assessing statistical power of SNPs for population structure and conservation studies. Molecular Ecology Resources 9:66-73.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, USA.

Noad, M.J., R.A. Dunlop, D. Paton and D.H. Cato. 2011. Absolute and relative abundance estimates of Australian east coast humpback whales (*Megaptera novaeangliae*). Journal of Cetacean Research and Management (special issue 3):243-252.

Olavarria, C., M. Anderson, D. Paton, D. Burns, M. Brasseur, C. Garrigue, N. Hauser, et al. 2006. Eastern Australia humpback whale genetic diversity and their relationship with Breeding Stocks D, E, F and G. Paper SC/58/SH25 presented at the IWC Scientific Committee, 2006 (unpublished).

Olavarría, C., C.S. Baker, C. Garrigue, M. Poole, N. Hauser, S. Caballero, L. Flórez-González, et al. 2007. Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. Marine Ecology Progress Series 330:257-268.

Olsen, J.B., P. Bentzen, M.A. Banks, J.B. Shaklee and S. Young. 2000. Microsatellites reveal population identity of individual pink salmon to allow supportive breeding of a population at risk of extinction. Transactions of the American Fisheries Society 129:232-242.

Pastene, L.A., M. Goto, N. Kanda, T. Kitakado and P. Palsboll. 2011. Preliminary mitochondrial DNA analysis of low and high latitude humpback whales of Stocks D, E and F. Paper SC/63/SH9 presented at the IWC Scientific Committee, Tromso, Norway.

Paton, D., A. Oosterman, M. Whicker and I. Kenny. 2006. Preliminary assessment of sighting survey data of humpback whales, Norfolk Island, Australia. Paper SC/A06/HW36 submitted to the IWC southern hemisphere humpback workshop, Hobart, April 2006.

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Peakall, R. and P.E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics 28:2537-2539.

Peakall, R., P.E. Smouse and D.R. Huff. 1995. Evolutionary implications of allozyme and RAPD Variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. Molecular Ecology 4:135-147.

Pella, J.J. and G.B. Milner. 1987. Use of genetic marks in stock composition analysis *in* N. Ryman and F. Utter, ed. Population genetics and fishery management. University of Washington Press, Seattle, USA.

Piry, S., A. Alapetite, J.M. Cornuet, D. Paetkau, L. Baudouin and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity 95:536-539.

Rannala, B. and J.L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the United States of America 94:9197-9201.

Rock, J., Pastene, L.A., Kaufman, G., Forestell, P., Matsuoka, K. and Allen, J. 2006. A note on East Australia Group V Stock humpback whale movement between feeding and breeding areas based on photo-identification. Journal of Cetacean Research and Management 8:301-305.

Rosenbaum, H.C., C. Pomilla, M. Mendez, M.S. Leslie, P.B. Best, K.P. Findlay, G. Minton, et al. 2009. Population Structure of Humpback Whales from Their Breeding Grounds in the South Atlantic and Indian Oceans. PLoS ONE 4:e7318.

Schmitt, N.T., M.C. Double, C.S. Baker, D. Steel, K.C.S. Jenner, M.-N.M. Jenner, D. Paton, et al. In Press. Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations. Marine Mammal Science.

Selkoe, K.A. and R.J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9:615-629.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.

Smith, C.T., W.D. Templin, J.E. Seeb and U.W. Seeb. 2005. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of US and Canadian Chinook salmon caught in Yukon River fisheries. North American Journal of Fisheries Management 25:944-953.

Steel, D., C. Garrigue, M. Poole, N. Hauser, C. Olavarria, L. Florez-Gonzalez, R. Constantine, et al. 2008. Migratory connections between humpback whales from South Pacific breeding grounds and Antarctic feeding areas based on genotype matching. Paper SC/60/SH13 presented at the IWC Scientific Committee, Santiago, Chile.

Steel, D., N.T. Schmitt, M. Anderson, D. Burns, S. Childerhouse, R. Constantine, T. Franklin, et al. 2011. Initial genotype matching of humpback whales from the 2010 Australia/New Zealand Antarctic Whale Expedition (Area V) to Australia and the South Pacific. Paper SC/63/SH10 presented at the IWC Scientific Committee, Trömso, Norway.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in Ecology & Evolution 15:199-203.

Topchy, A., K. Scribner and W. Punch. 2004. Accuracy-driven loci selection and assignment of individuals. Molecular Ecology Notes 4:798-800.

Utter, F. and N. Ryman. 1993. Genetic markers and mixed stock fisheries. Fisheries 18:11-21.

VanDeHey, J.A., B.L. Sloss, P.J. Peeters and T.M. Sutton. 2010. Determining the Efficacy of Microsatellite DNA-Based Mixed Stock Analysis of Lake Michigan's Lake Whitefish Commercial Fishery. Journal of Great Lakes Research 36:52-58.

Waples, R.S., G.A. Winans, F.M. Utter and C. Mahnken. 1993. Genetic approaches to the management of pacific salmon. Fish 15:12-19.

Weir, B.S. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Zerbini, A.N., E. Ward, M.H. Engel, A. Andriolo and P.G. Kinas. 2006. A Bayesian assessment of the conservation status of humpback whales (Megaptera novaeangliae) in the western South Atlantic Ocean (Breeding stock A). 25pp. Paper SC/58/SH2 presented to the IWC Scientific Committee, 2006 (unpublished).

Table 1. Samples from individual whales (assumed from unique genotypes) genotyped at ten microsatellite loci and sequenced at the mtDNA control region for Antarctic feeding Area V and the six source populations.
M = males, F = females, ? = unknown

Region	Sampling period	*N*							
		Microsatellites	Sex			mtDNA	Sex		
			M	F	?		M	F	?
Antarctica Area V	1999-2010	**62**	31	31		**62**	31	31	
western Australia									
Exmouth	2007	204	116	88		189	107	82	
eastern Australia		190	127	62	1	180	141	59	
Eden	2008	61	47	14		57	44	13	
Tasmania	2006-2008	70	34	36		66	31	35	
*Evans Head**	2009	59	46	12	1	57	66	11	
Total Australia		**394**	**243**	**150**	**1**	**369**	**248**	**141**	
New Caledonia	1999-2005	310	172	123	15	299	170	121	8
Tonga	1999-2005	298	196	97	5	292	193	95	4
Cook Islands	1999-2005	95	47	42	6	91	46	42	3
French Polynesia	1999-2007	200	100	83	17	190	93	81	16
Total Oceania		**903**	515	345	43	**872**	502	339	31
Total		**1359**	789	526	44	**1303**	781	511	31

*Evans Head samples only used in the MSA of Antarctica Area V samples.

Table 2. Pairwise F_{ST} and D_{EST} values among source populations and Antarctic feeding Area V under (a) H1 and (b) H2 based on ten microsatellite loci. F_{ST} values are given below the diagonal and D_{EST} values are given above the diagonal. Significant P-values ($P < 0.05$ after sequential Bonferroni correction) for F_{ST}, based on statistical testing of 999 random permutations and for D_{EST}, based on 1000 bootstrap resamplings, are shown in bold.

Area V = Antarctic feeding Area V; WA = western Australia;
EA = eastern Australia; NC = New Caledonia; TG = Tonga;
CI = Cook Islands; FP = French Polynesia

(a) H1

Popn.	Area V	WA	EA	NC	TG	CI	FP
Area V		**0.028**	0.000	0.004	**0.019**	0.020	**0.035**
WA	**0.006**		**0.031**	**0.024**	**0.023**	**0.026**	**0.026**
EA	0.000	**0.005**		**0.014**	**0.023**	**0.020**	**0.023**
NC	0.002	**0.006**	**0.003**		**0.009**	0.008	**0.022**
TG	**0.016**	**0.014**	**0.015**	**0.013**		0.000	**0.007**
CI	**0.007**	**0.010**	**0.009**	**0.004**	**0.013**		0.000
FP	**0.008**	**0.005**	**0.005**	**0.005**	**0.009**	**0.004**	

(b) H2

Popn.	Area V	WA	EA	Oceania
Area V		**0.028**	0.000	**0.013**
WA	**0.006**		**0.031**	**0.021**
EA	0.000	**0.005**		**0.016**
Oceania	**0.005**	**0.005**	**0.005**	

Table 3. Pairwise F_{ST} and Φ_{ST} values among source populations and Antarctic feeding Area V under (a) H1 and (b) H2 based on mitochondrial DNA control region sequences. F_{ST} values are given below the diagonal and Φ_{ST} values are given above the diagonal. Significant P-values ($P < 0.05$ after sequential Bonferroni correction) based on statistical testing of 999 random permutations are shown in bold.

Area V = Antarctic feeding Area V; WA = western Australia;
EA = eastern Australia; NC = New Caledonia; TG = Tonga;
CI = Cook Islands; FP = French Polynesia

(a) H1

Popn.	Area V	WA	EA	NC	TG	CI	FP
Area V		**0.024**	0.001	0.003	0.005	0.018	**0.047**
WA	**0.016**		**0.042**	**0.017**	**0.022**	**0.045**	**0.082**
EA	0.002	**0.015**		**0.021**	0.012	0.012	**0.036**
NC	**0.005**	**0.016**	**0.010**		0.005	**0.019**	**0.042**
TG	**0.015**	**0.016**	**0.010**	**0.009**		0.007	**0.035**
CI	**0.033**	**0.031**	**0.025**	**0.030**	**0.010**		0.005
FP	**0.037**	**0.034**	**0.031**	**0.031**	**0.017**	0.003	

(b) H2

Popn.	Area V	WA	EA	Oceania
Area V		**0.024**	0.001	0.006
WA	**0.016**		**0.042**	**0.027**
EA	0.002	**0.015**		**0.013**
Oceania	**0.010**	**0.015**	**0.010**	

Table 4. Individual assignment test results using the maximum likelihood clasification in the program MLE (Topchy *et al* . 2004). Individuals from the three genetic datasets (mtDNA, microsatellites and mDNA + microsatellites) were re-assigned to the most likely population in mixtures likely to occur on the Antarctic feeding grounds by calculating the posterior probability for each classification.

WA = western Australia; EA = eastern Australia; NC = New Caledonia; TG = Tonga; CI = Cook Islands; FP = French Polynesia

		mtDNA	microsatellites	mtDNA + microsatellites
Mixture	Population	% individuals re-assigned correctly	% individuals re-assigned correctly	% individuals re-assigned correctly
western Australia and	WA	48.2	33.8	50.5
eastern Australia	EA	28.4	26.0	38.9
	total	**40.3**	**30.7**	**45.9**
eastern Australia and	EA	21.9	11.5	27.5
Oceania	Oceania	23.0	28.1	45.9
	total	**22.9**	**25.7**	**43.3**
eastern Australia, New Caledonia,	EA	6.5	2.3	16.0
Tonga, Cook Islands and	NC	0.3	0.0	5.8
French Polynesia	TG	0.0	0.3	2.0
	CI	0.0	0.0	0.0
	FP	3.2	3.0	7.5
	total	**1.5**	**1.0**	**5.8**

Appendix 1. Genetic diversity in humpback whales sampled from Antarctic feeding Area V, the six source populations, and the Evan's Head data from eastern Australia genotyped at ten loci. *N* = number of genotyped indviduals per locus, *Na* = number of alleles, *Ho* = observed heterozygosity, *He* = expected heterozygosity and HW = deviation from Hardy-Weinberg equilibrium (p-value); significant at P < 0.05 after the adjustment for multiple comparison with sequential Bonferroni test (Rice 1989) (Standard errors in parentheses).

	Antarctic Area V					western Australia					eastern Australia					Evan's Head (eastern Australia)				
Locus	*N*	*Na*	*Ho*	*He*	HW	*N*	*Na*	*Ho*	*He*	HW	*N*	*Na*	*Ho*	*He*	HW	*N*	*Na*	*Ho*	*He*	HW
Ev14	62	9	0.758	0.723	0.426	203	8	0.754	0.778	0.458	131	9	0.725	0.748	0.742	59	11	0.729	0.757	0.367
Ev37	62	16	0.935	0.909	0.372	202	19	0.931	0.904	0.322	131	19	0.916	0.913	0.360	59	15	0.898	0.910	0.043
Ev96	62	13	0.823	0.852	0.049	202	12	0.876	0.869	0.688	131	13	0.863	0.848	0.867	59	12	0.915	0.864	0.641
GATA417	58	15	0.862	0.883	0.420	203	15	0.911	0.903	0.741	131	15	0.870	0.890	0.631	59	13	0.898	0.877	0.043
GT211	62	9	0.758	0.787	0.868	203	10	0.803	0.836	0.033	130	10	0.785	0.820	0.675	59	9	0.644	0.849	0.024
GT23	62	8	0.823	0.759	0.492	204	9	0.838	0.821	0.618	131	9	0.763	0.797	0.185	59	8	0.797	0.792	0.410
rw4-10	59	10	0.932	0.835	0.819	203	12	0.877	0.854	0.798	131	12	0.786	0.831	0.582	59	10	0.847	0.825	0.123
Ev1	62	4	0.452	0.565	0.006	203	4	0.567	0.526	0.428	130	4	0.523	0.552	0.429	59	4	0.694	0.581	0.434
Ev94	62	10	0.823	0.801	0.182	202	9	0.827	0.809	0.357	130	9	0.792	0.807	0.760	59	8	0.780	0.831	0.081
GT575	61	14	0.803	0.783	0.703	203	16	0.788	0.804	0.292	130	14	0.815	0.811	0.824	59	11	0.847	0.819	0.578
all loci	61.2 (*0.5*)	10.8 (*1.2*)	0.80 (*0.04*)	0.79 (*0.03*)	0.43 (*0.10*)	202.8 (*0.20*)	11.4 (*1.4*)	0.82 (*0.03*)	0.81 (*0.03*)	0.47 (*0.08*)	130.6 (*0.2*)	11.4 (*1.3*)	0.78 (*0.03*)	0.80 (*0.03*)	0.61 (*0.07*)	59 (*0.0*)	10.1 (*2.9*)	0.81 (*0.09*)	0.81 (*0.09*)	0.27 (*0.08*)

* Of the 13,590 genotypes in total we failed to genotype 384, with only two of 1,359 individuals missing genotypes from three or more loci and no individuals missing data from more than four of ten loci.

New Caledonia					Tonga					Cook Islands					French Polynesia				
N	*Na*	*Ho*	*He*	HW	*N*	*Na*	*Ho*	*He*	HW	*N*	*Na*	*Ho*	*He*	HW	*N*	*Na*	*Ho*	*He*	HW
309	9	0.725	0.729	0.558	188	10	0.766	0.773	0.344	94	9	0.830	0.798	0.120	189	9	0.735	0.761	0.014
292	21	0.932	0.925	0.910	288	19	0.913	0.921	0.025	90	17	0.867	0.921	0.011	199	19	0.920	0.909	0.761
310	13	0.910	0.880	0.375	292	12	0.839	0.878	0.303	87	12	0.770	0.864	0.039	183	12	0.869	0.867	0.309
275	18	0.931	0.904	0.378	294	21	0.925	0.909	0.935	75	17	0.920	0.898	0.790	197	21	0.873	0.905	0.590
310	10	0.852	0.831	0.425	290	10	0.824	0.825	0.637	94	10	0.840	0.816	0.869	199	9	0.804	0.823	0.437
309	9	0.822	0.790	0.883	292	9	0.818	0.795	0.226	95	9	0.811	0.778	0.771	196	9	0.806	0.811	0.265
307	11	0.834	0.844	0.144	292	13	0.795	0.824	0.289	77	12	0.805	0.813	0.938	196	11	0.801	0.822	0.941
300	4	0.500	0.506	0.057	288	4	0.479	0.480	0.244	94	4	0.511	0.501	0.624	193	4	0.430	0.455	0.510
309	10	0.848	0.807	0.461	293	9	0.775	0.805	0.194	90	9	0.711	0.813	0.020	192	9	0.802	0.803	0.034
308	15	0.828	0.814	0.417	292	15	0.815	0.833	0.391	95	12	0.758	0.817	0.300	197	14	0.838	0.806	0.688
302.9 (*3.6*)	12.0 (*1.6*)	0.82 (*0.04*)	0.80 (*0.04*)	0.46 (*0.09*)	280.9 (*10.3*)	12.2 (*1.6*)	0.80 (*0.04*)	0.80 (*0.04*)	0.36 (*0.08*)	89.1 (*2.3*)	11.1 (*1.2*)	0.78 (*0.04*)	0.80 (*0.04*)	0.45 (*0.12*)	194.1 (*1.6*)	11.7 (*1.6*)	0.79 (*0.04*)	0.80 (*0.04*)	0.46 (*0.10*)

Appendix 2. Variability in the mtDNA control region of humpback whales sampled from Antarctic feeding Area V, the source populations, and Evan's Head (eastern Australia).
(h = haplotype diversity and π = nucleotide diversity)

Area V = Antarctic feeding Area V; WA = western Australia;
EA = eastern Australia; NC = New Caledonia; TG = Tonga;
CI = Cook Islands; FP = French Polynesia

	Region/Popn	No. of haplotypes	No. of unique haplotypes	$h \pm$ SD	$\pi \pm$ SD
Antarctica	Area V	30	3	0.969±0.008	0.141±0.072
	WA	56	23	0.971±0.004	0.132±0.067
	EA	40	4	0.966±0.005	0.126±0.065
	NC	64	9	0.973±0.002	0.138±0.070
	TG	53	4	0.964±0.003	0.136±0.069
	CI	29	1	0.925±0.016	0.125±0.065
	FP	30	3	0.919±0.010	0.114±0.059
	Oceania	80	30	0.986±0.009	0.133±0.067
Evan's Head	EA	27	1	0.946±0.018	0.199±0.103
Total		114	48	0.973±0.002	0.133±0.067

Figure 1. Australian and Oceania humpback whale breeding/migratory populations and their associated feeding areas (Area IV, V, VI and I) divided into 'Pure Stock' and 'Mixing' areas (IWC 2009). We also include the number of individuals sampled from each region (*N*). 'Oceania' combines New Caledonia, Tonga and French Polynesia in a population structure hypothesis proposed for the Comprehensive Assessment of Southern Hemisphere humpback whales (IWC 2006). WA = western Australia; EA = eastern Australia.

Figure 2. Results of 100% simulations for humpback whale populations under a) H1 and b) H2. The program SPAM was used to simulate a 'pure stock' (*N* = 40) consisting of 100% of each source population. We then assessed the ability of mixed stock analysis to correctly estimate the reapportionment of each population using mitochondrial DNA (mtDNA) control region sequences, ten microsatellite loci, and mtDNA + microsatellites combined. We expect to see 100% of the correct population. Error bars represent 95% confidence intervals.

WA = western Australia; EA = eastern Australia; NC = New Caledonia; TG = Tonga; CI = Cook Islands; FP = French Polynesia

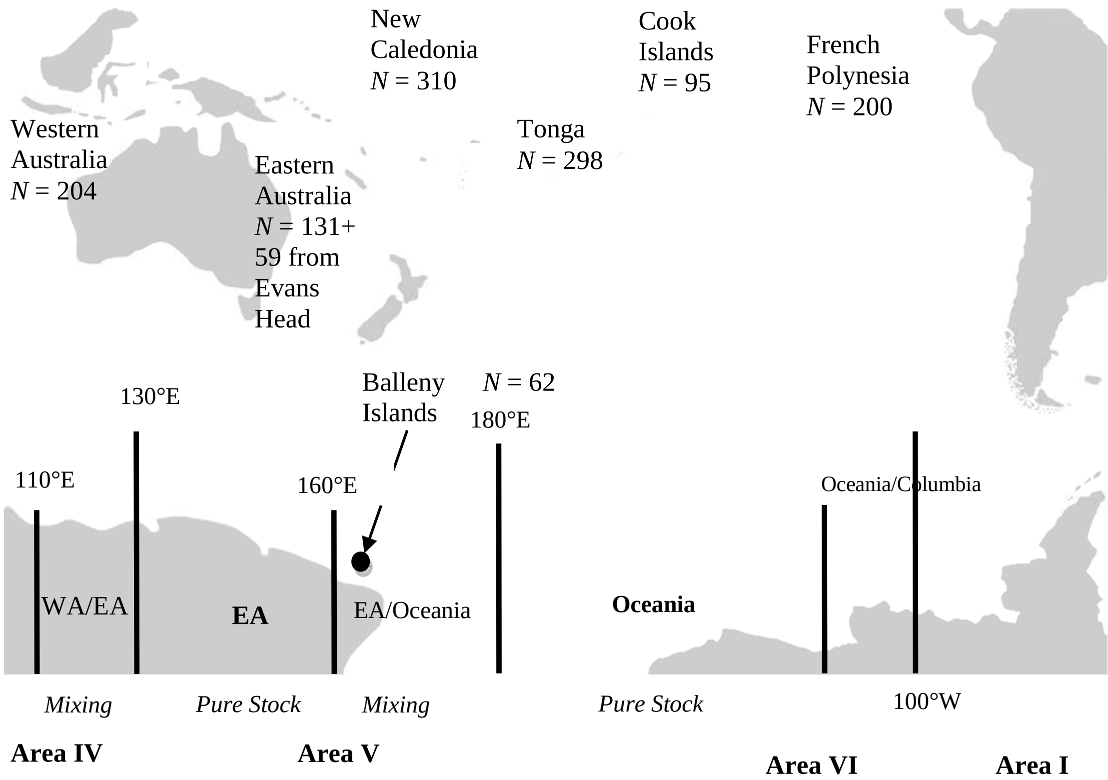
Figure 3. The effect of increasing population sample size on the genetic distinctiveness of populations under H2 for the combined mtDNA + microsatellite dataset using the 100% simulation analysis in SPAM and a simulation sample size of *N* = 40. Error bars represent 95% confidence intervals.

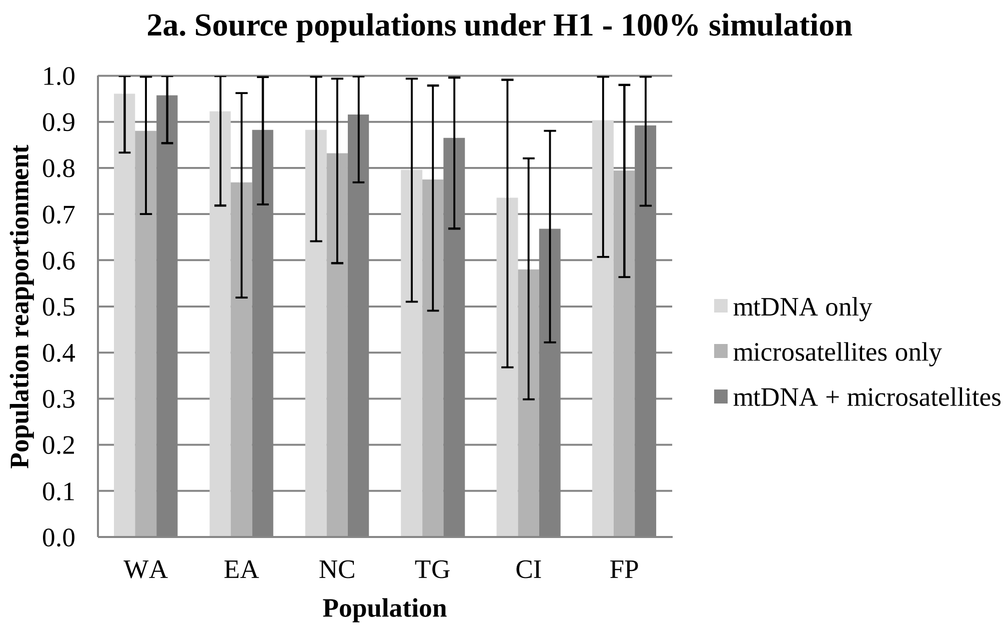
Figure 4. Mean reapportionment for a) western Australian humpback whales in a western Australia/eastern Australia mix, and b) Oceania humpback whales in a

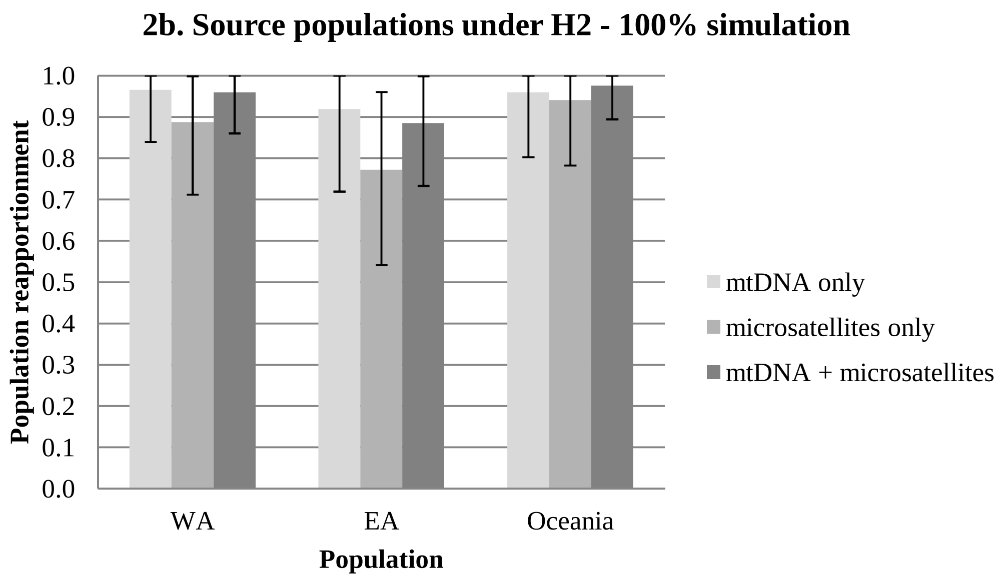
Oceania/eastern Australia mix, simulated as a function of feeding ground (mixture) sample size using the 100% simulation feature in the program SPAM and the combined microsatellite and mtDNA control region dataset. Error bars represent 95% confidence intervals. The horizontal line represents where the mean correct assignment begins to stabilize. In both figures the dashed circle encompasses the feeding ground sample size where the mean reapportionment matches the horizontal line value and confidence intervals begin to stabilize (*N* = 160).

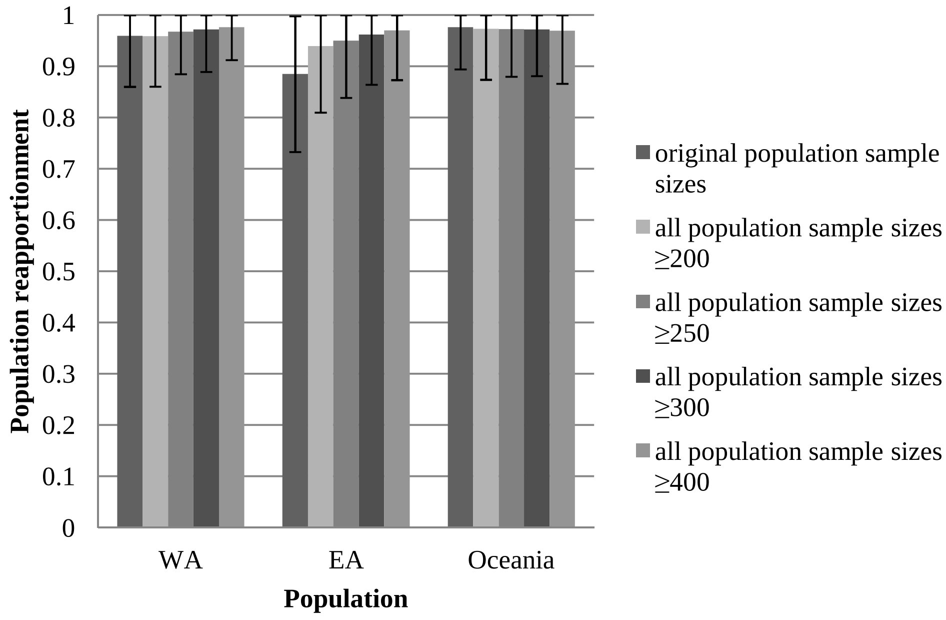
Figure 5. Results of simulated mixture analyses for two humpback whale feeding ground mixtures; eastern Australia/Oceania and western Australia/eastern Australia. We used the program SPAM to simulate mixtures with expected proportional contributions varying from 0.0 to 1.0. We then performed mixed stock analysis, calculated maximum likelihood estimates of the proportional contributions in each simulated mixture, and compared estimated values to the true expected proportions using the combined microsatellite + mtDNA control region datasets. Figures a) and b) used original population sample sizes and a mixture sample size of *N* = 40; Figures c) and d) used population sample sizes ≥ 200 and a mixture sample size of *N* = 40; Figures e) and f) used population sample sizes ≥ 200 and a mixture sample of *N* = 160.

Figure 6. SPAM maximum likelihood estimates of humpback whale feeding Area V composition (*N* = 62) under H1 (b and d) and H2 (a and c) using all three genetic marker sets. Error bars represent jackknife standard errors. Figures a and b show estimates using source data and figures c and d show estimates using source data with eastern Australia supplemented by 59 Evan’s Head samples.

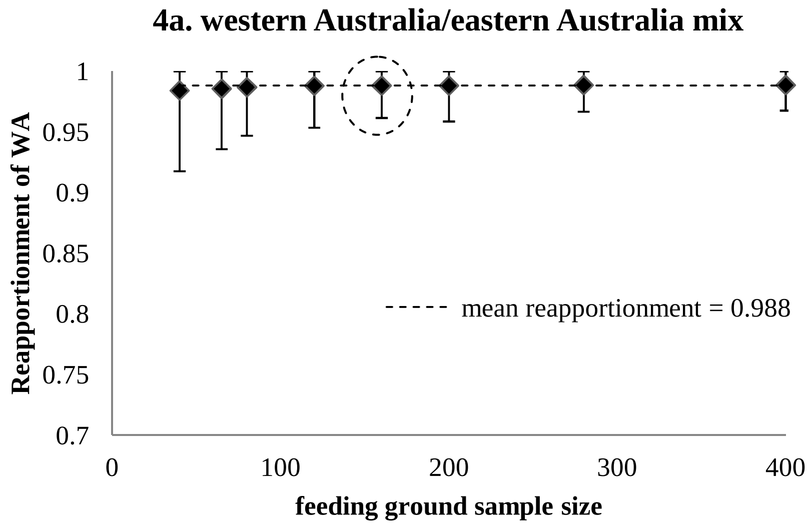

New
Caledonia
N = 310
Cook
Islands
N = 95
French
Polynesia
N = 200
Western
Australia
N = 204
Tonga
N = 298
Eastern
Australia
N = 131+
59 from
Evans
Head
Balleny
Islands
N = 62
130°E
180°E
110°E
160°E
Oceania/Columbia
WA/EA
EA
EA/Oceania
Oceania
Mixing
Pure Stock
Mixing
Pure Stock
100°W
Area IV
Area V
Area VI
Area I

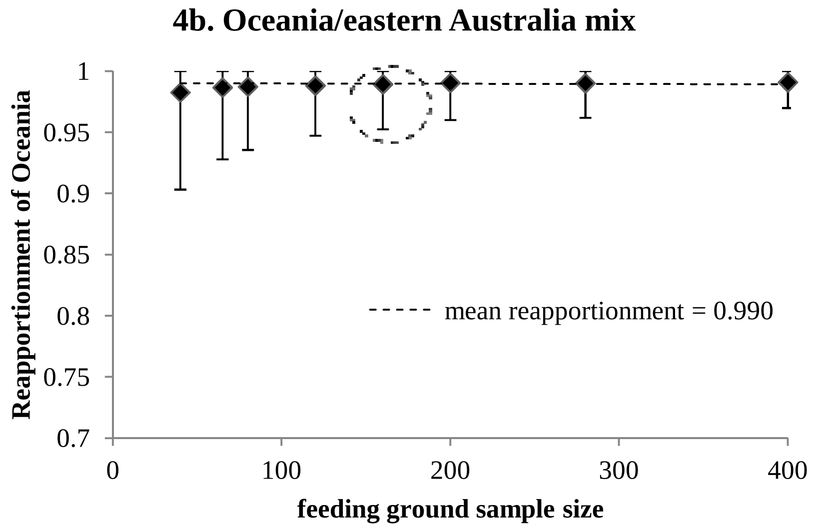

2a. Source populations under H1 - 100% simulation
Population reapportionment
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WA
EA
NC
TG
CI
FP
Population
mtDNA only
microsatellites only
mtDNA + microsatellites


2b. Source populations under H2 - 100% simulation
Population reapportionment
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WA
EA
Oceania
Population
mtDNA only
microsatellites only
mtDNA + microsatellites

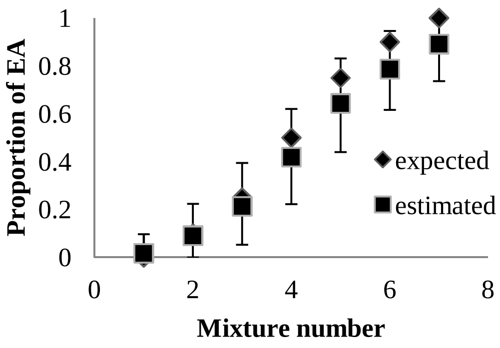


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Population reapportionment
WA
EA
Oceania
Population
original population sample sizes
all population sample sizes ≥200
all population sample sizes ≥250
all population sample sizes ≥300
all population sample sizes ≥400

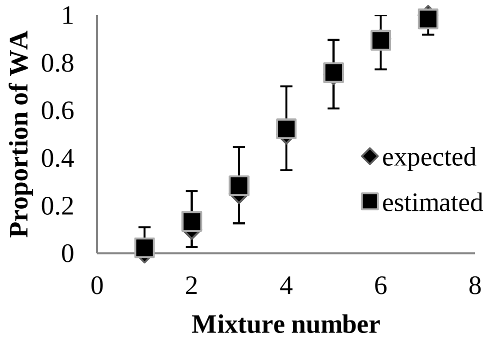

4a. western Australia/eastern Australia mix
Reapportionment of WA
1
0.95
0.9
0.85
0.8
0.75
0.7
0
100
200
300
400
mean reapportionment = 0.988
feeding ground sample size


4b. Oceania/eastern Australia mix
Reapportionment of Oceania
1
0.95
0.9
0.85
0.8
0.75
0.7
0
100
200
300
400
mean reapportionment = 0.990
feeding ground sample size

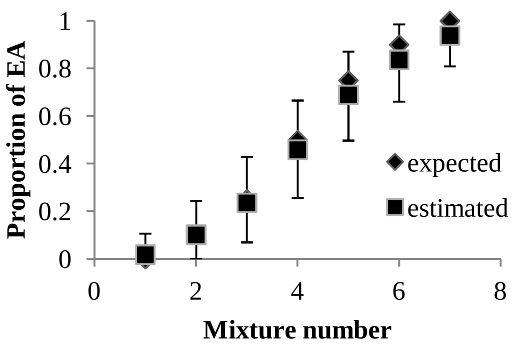
5a. eastern Australia/Oceania mix N = 40 (original source sample sizes)



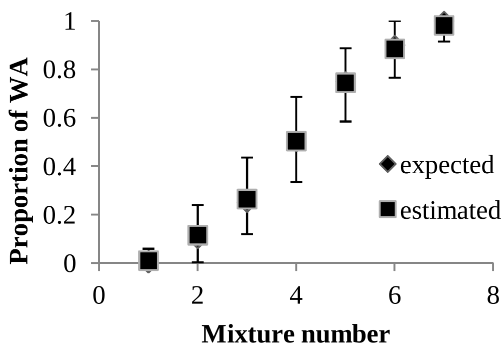
5b. western Australia/eastern Australia mix N = 40 (original source sample sizes)



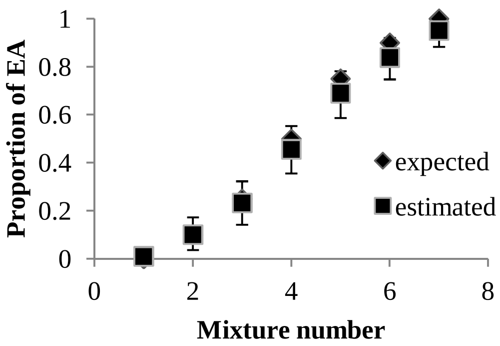
5c. eastern Australia/Oceania mix N = 40 (EA sample size increased to 200)



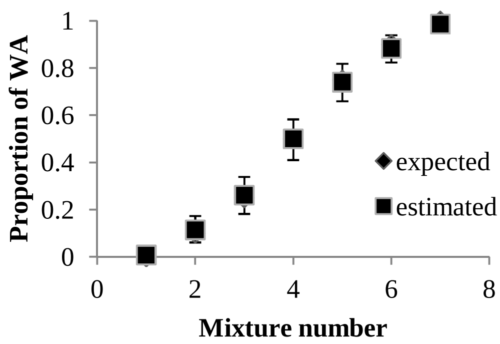
5d. western Australia/eastern Australia mix N = 40 (EA sample size increased to 200)

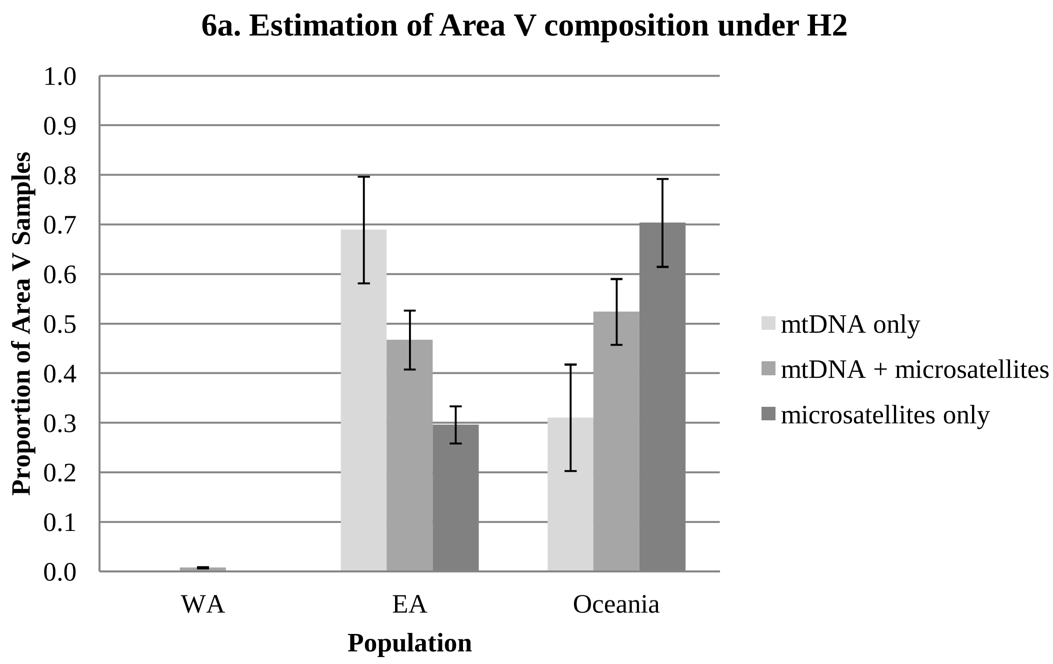


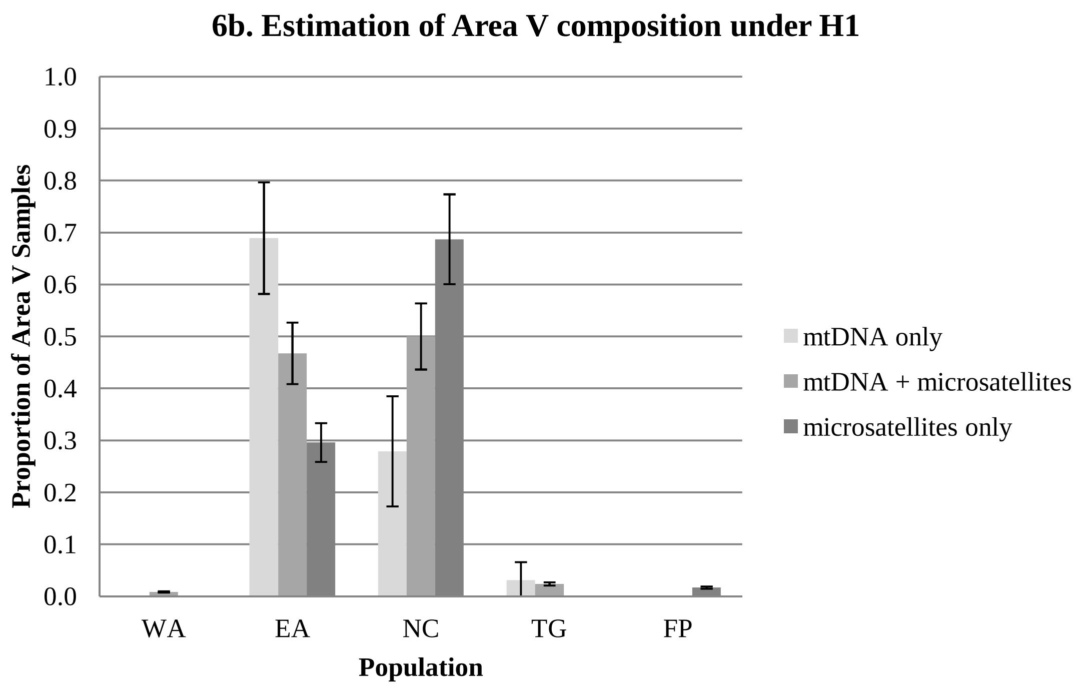
5e. eastern Australia/Oceania mix N = 160 (EA sample size increased to 200)

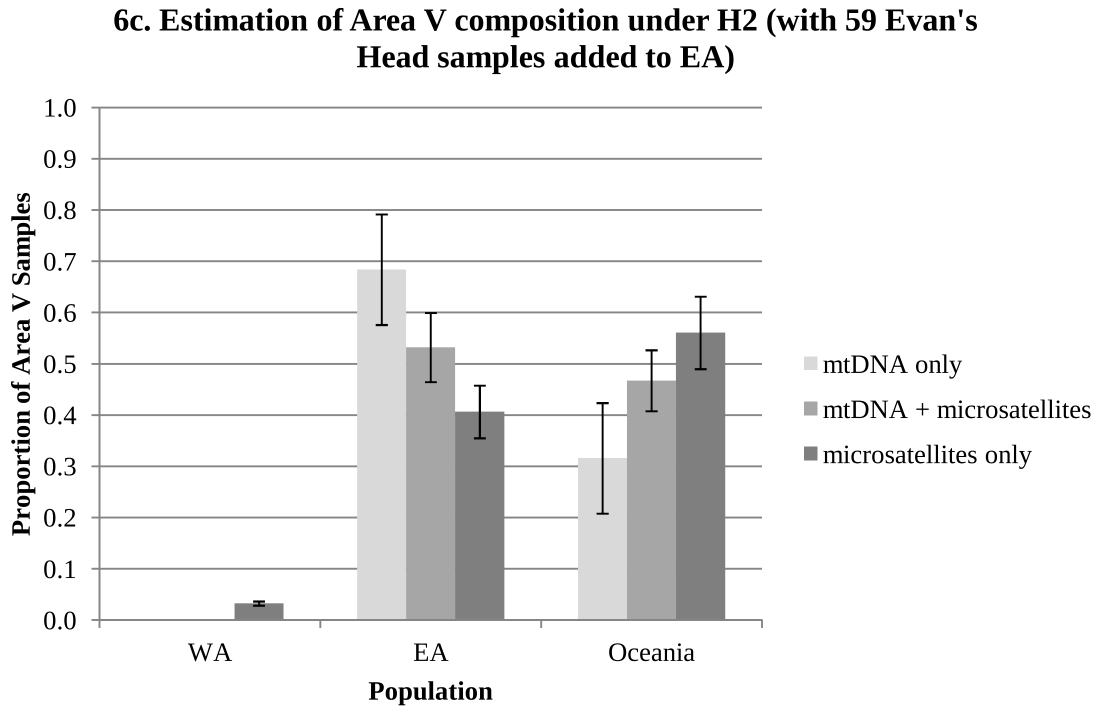


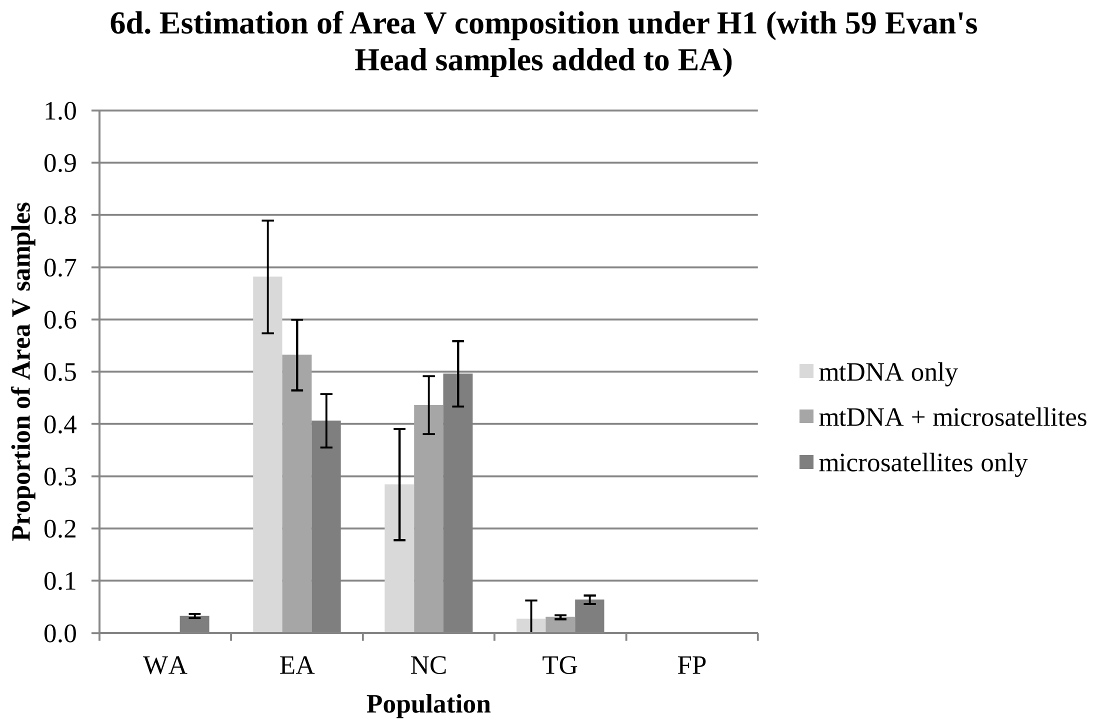
5f. western Australia/eastern Australia mix N = 160 (EA sample size increased to 200)




6a. Estimation of Area V composition under H2
Proportion of Area V Samples
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WA
EA
Oceania
Population
mtDNA only
mtDNA + microsatellites
microsatellites only


6b. Estimation of Area V composition under H1
Proportion of Area V Samples
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WA
EA
NC
TG
FP
Population
mtDNA only
mtDNA + microsatellites
microsatellites only


6c. Estimation of Area V composition under H2 (with 59 Evan's Head samples added to EA)
Proportion of Area V Samples
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WA
EA
Oceania
Population
mtDNA only
mtDNA + microsatellites
microsatellites only


6d. Estimation of Area V composition under H1 (with 59 Evan's Head samples added to EA)
Proportion of Area V samples
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WA
EA
NC
TG
FP
Population
mtDNA only
mtDNA + microsatellites
microsatellites only

Chapter 3: Re-assessment of the genetic evidence for sex specific migratory route choice in eastern Australian humpback whales (*Megaptera novaeangliae*)

NATALIE T. SCHMITT[1,2], MICHAEL C. DOUBLE[1], SIMON N. JARMAN[3], ANDREA M. POLANOWSKI[1,3], ROSEMARY GALES[4], ROD PEAKALL[2]

[1]Australian Marine Mammal Centre, Australian Antarctic Division, Kingston, Tasmania 7050, Australia

[2]Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra ACT 0200, Australia

[3]Australian Antarctic Division, Kingston, Tasmania 7050, Australia

[4]DPIPWE, 134 Macquarie St, Hobart, Tasmania 7000, Australia

ABSTRACT

While the coarse-scale spatial and temporal patterns of humpback whale seasonal movements between low latitude breeding grounds and high latitude feeding grounds are broadly understood, the assumption that the general pattern of migration for both sexes is similar is still under question. One study by Valsecchi et al. (2010) has challenged this prevailing view for the humpback whales that migrate along the east coast of Australia based on a finding of significant mtDNA nucleotide differentiation between northbound males and females sampled off Stradbroke Island in 1992. They suggested that sex-specific patterns of migration are operating such that when sampled at the same location, males and females represent different populations.

Using sex-specific and temporal mtDNA data presented in this previous study together with 180 additional samples from two new sampling regions along the eastern Australian migratory corridor (Evan's Head and south-eastern Australia), we reassessed the evidence for sex specific migratory route choice. Specifically we determined whether the patterns of haplotype sharing, haplotype frequency and haplotype and nucleotide differentiation are consistent with the null hypothesis that males and females drawn from the same population will be genetically similar.

In our initial analysis of the mtDNA data, we found significant structure among all sampling locations at the haplotype level, irrespective of sex or migratory direction, and for whales migrating north towards the breeding grounds. When samples were partitioned by sex and sampling location, we recovered the same intriguing finding as the previous study of significant structure between Stradbroke Island males and females at the nucleotide level, both overall and for northbound individuals, but this was not repeated in the samples off Evan's Head and south-eastern Australia. Across all three

sampling locations, regardless of migratory direction, we found no significant differences in the patterns of haplotype differentiation, haplotype sharing and haplotype frequency between the sexes. Overall, our results raise doubts about whether eastern Australian males and females are utilizing different migratory routes, as well as highlight other potential anomalies.

We discuss possible explanations for these findings, including sampling, mixing with neighbouring populations, demographic consequences of historical overexploitation, and social and age class structuring during migration.

KEY WORDS: Humpback whales, mtDNA control region, breeding grounds, eastern Australia

*Email: natalie.schmitt@aad.gov.au, natschmitt@hotmail.com

INTRODUCTION

Long-distance, seasonal migration is displayed by all major animal groups but arguably the most commonly known examples are from species of birds and marine mammals (Dingle & Drake 2007). In both hemispheres, many species of birds migrate hundreds, even thousands of kilometres from tropical to temperate regions to breed in spring and perform the return journey the following autumn (e.g. Cox 1985). Similarly it has been shown that many marine species make large-scale latitudinal migrations which may also be driven by mating opportunities, as well as the abundance of suitable prey, thermal habitat preferences and ocean processes (Block et al. 2011). The long-distance seasonal migration of the humpback whale *Megaptera novaeangliae* is one of the most intensely studied migrations of all marine species. Occurring in all ocean basins, these whales forage in productive, high-latitude areas in the summer before undertaking extensive movements to low-latitude coastal waters, where mating and calving occur during winter and spring (Dawbin 1966).

With the development of satellite and geolocation tags, tracking studies of long-distance, seasonal migration behaviour is now commonplace among migratory species, revealing different strategies during migration dependent on sex, age or condition. For example, in many bird species, the males depart their winter habitats first to establish a breeding territory on the summer breeding grounds prior to the later arrival of females (e.g. Stanley et al. 2012). Such studies have also discovered that the route of migration can vary greatly; even the same individual may change its migration route between years based on its own condition or climatic variables (e.g. Yosef et al. 2006, Purcell & Brodin 2007, Senapathi et al. 2011, Gronroos et al. 2012). However, despite this large and rapidly growing body of literature, where migratory behaviour can be strongly influenced by an individual's life-history, there are very few examples where the actual migratory route differs substantially between sexes within a population. It was therefore surprising when Valsecchi et al. (2010) raised this possibility in humpback whales. Their study found significant mtDNA nucleotide differentiation between northbound males and female whales sampled from the migratory corridor off Stradbroke Island (Queensland) in 1992. To explain this result the authors suggested that sex-specific patterns of migration are operating in humpback whales that migrate along eastern Australia. Further, based on behavioural evidence, they argued their best supported hypothesis was one where eastern Australia males migrate northward to the breeding ground elsewhere in the western South Pacific, whereas females migrate along the east coast of Australia. The males that migrate north along eastern Australia are thought to belong to a separate western South Pacific population.

This intriguing result and hypothesis is not implausible given the unique environment of the western and central South Pacific Ocean for humpback whales. Thousands of islands and reefs create a band of suitable wintering habitat for a number of breeding populations in close proximity to each other, where longitudinal movements between adjacent populations or feeding areas are not constrained by large land masses (Garrigue et al. 2000). Valsecchi et al. (2010) suggested that such longitudinal movements may be adaptive if males can opportunistically visit many breeding sites at the 'peak' of female receptivity within one migratory season.

The Valsecchi et al. (2010) result demands further assessment for several reasons. Firstly there have been several behavioural, genetic and tagging studies on western South Pacific humpback whales that have revealed temporal segregation along migratory routes according to age, sex and reproductive state (Chittleborough 1965, Dawbin 1966, Brown & Corkeron 1995, Brown et al. 1995, Craig & Herman 1997, Smith et al. 2012), but have not indicated sex-specific migratory route choice. Secondly, it is thought baleen whale calves develop a “cultural memory” of particular migratory routes from their mother prior to weaning thus producing maternally directed fidelity to feeding and breeding areas (Clapham & Mayo 1987, Clapham & Seipt 1991, Clapham et al. 1993, Clapham 1996); the Valsecchi et al. (2010) result challenges this view, suggesting there may be male-specific migratory routes. Finally, if indeed there is sex-specific migratory route choice in the western South Pacific, then this would raise doubt over the validity and interpretation of the important and influential population abundance estimates for eastern Australian humpback whales derived from surveys off Stradbroke Island (Noad et al. 2005, Noad et al. 2011). These data are not only used to assess the population trend of the eastern Australian population but also to derive biological parameters used in population, recovery and even whaling management models developed by the International Whaling Commission (e.g. Best 2001, Clapham 2001, IWC 2001, Baker & Clapham 2004, Jackson et al. 2008).

In this study we reassess the genetic evidence of sex specific migratory route choice in western South Pacific humpback whales using the data presented by Valsecchi et al. (2010) together with 180 additional samples from two new sampling regions along the eastern Australian migratory corridor.

MATERIALS AND METHODS

Samples

Our data set consisted of 315 mtDNA sequences derived from humpback whale skin biopsy samples from the eastern Australian migratory corridor (Table 1). This included

135 sequences (91 males and 44 females) from samples collected off Stradbroke Island in 1992 throughout the annual migration, published by Valsecchi et al. (2010) and 180 sequences (121 males and 59 females) from sampling efforts between 2007 and 2009 around Evan's Head, NSW and south-eastern Australia, which included Eden, NSW, and eastern Tasmania (see Schmitt et al. In Press). See table 1 for a breakdown of samples according to sex and migratory direction as well as the sample collection dates. All eastern Australian samples were collected during the peaks of both the northern and southern migrations (Paterson et al. 1994), with the Evan's Head sample collection overlapping the Valsecchi et al. (2010) sampling period for the northern migration. Further details on the sample collection and sex determination can be found in a previous paper (Schmitt et al. In Press).

mtDNA sequence

For all samples, our analysis was based on a 470bp consensus region starting at position six of the reference humpback whale mitochondrial control region sequence (GenBank X72202: see Baker & Medrano-Gonzalez 2002, Olavarría et al. 2007). The Stradbroke Island sequences were extracted from Valsecchi et al. (2010) based on their haplotype identity referenced from Olavarria et al. (2007).

Statistical Analysis

By virtue of its maternal inheritance, interbreeding populations will share mtDNA haplotypes. Thus, in the absence of sex-biased dispersal, samples of males and females drawn from the same population will be genetically similar. For mtDNA there are several different, but complementary ways to compare the genetic similarity between the sexes. At the simplest level, the number of different haplotypes shared, relative to the total number of haplotypes, is a potentially informative measure. A comparison of haplotype frequencies provides another way to assess genetic similarity, while quantifying the extent of genetic differentiation offers a third approach. If adequate samples of males and females were drawn from the same interbreeding population, we would expect to find high levels of haplotype sharing, haplotype frequencies to be similar, and haplotype differentiation to be absent.

To test for the presence of sex specific migratory route choice among our eastern Australian samples we employed three complementary approaches using the software GenAlEx 6.5 (Peakall & Smouse 2006, 2012). For these analyses we partitioned the data in four different ways: 1. by sampling location, 2. by sampling location and migratory direction, 3. between sexes within sampling locations, 4. between sexes by migratory direction within sampling locations.

I. Analysis of Molecular Variance (AMOVA)

The Analysis of Molecular Variance (AMOVA) framework allows the hierarchical partitioning of genetic variation between groups and the estimation of the widely used *F*-statistics and/or their analogues. We implemented AMOVA analyses following Excoffier et al. (1992), Huff et al. (1993) and Peakall et al. (1995), to partition the genetic variance within and among groups with statistical tests by random permutation (999 permutations). We employed permutation testing using the 'specialised' permute option in GenAlEx. This option performs an additional permutation procedure where either individuals are shuffled among sexes within sampling locations (F_{PR} or Φ_{PR}), or the sexes are shuffled among sampling locations (F_{RT} or Φ_{RT}) to estimate probability. This permutation option was chosen due to the detection of significant 'regional' structure, or structure among sampling locations (see RESULTS I). We also performed pairwise analyses where appropriate, where samples were permuted between each pairwise contrast of interest.

Within GenAlEx, the data type and choice of distance calculation used as input for AMOVA led to related but different analyses. We report here results for both the haplotype and nucleotide level differentiation. For the haplotype analysis, the list of numerically coded haplotypes (1 for each individual) formed the basis of the input, with the pairwise genetic distance between the haplotypes being either 0 (same) or 1 (different). At the nucleotide level, genetic distance matrices were constructed using individual pairwise differences across the polymorphic nucleotide sites (following Excoffier et al. 1992). We use the notation F_{ST} for haplotype differentiation and Φ_{ST} for nucleotide differentiation (e.g. Weir & Cockerham 1984, Takahata & Palumbi 1985, Hudson et al. 1992).

II. Haplotype sharing

By virtue of maternal inheritance and the lack of recombination, haplotypes will be invariably shared with other maternally related individuals (Ebert and Peakall, 2009a). Haplotype sharing (or lack thereof) can thus provide important clues about the distinctiveness between groups, which complements and extends other genetic analysis, such as estimates of genetic differentiation. To facilitate this analysis we extended the capability of GenAlEx 6.5 to determine the number of haplotypes shared relative to the total, between sexes within sampling locations, and between sampling locations. To assess whether or not the number of haplotypes shared between the sexes was consistent with males and females being from the same genetic population, we employed permutational testing using procedures analogous to those used in AMOVA, where haplotypes were randomly permuted between each pairwise contrast of interest (999 permutations), as described above. For each permutation, the number of haplotypes shared was recomputed and that value compared with the observed value. By tallying the number of times the observed value was smaller or larger than the permuted value we were able to perform both one-tailed and two-tailed statistical tests of departure from the null hypothesis (with only the two-tailed test reported).

III. Haplotype frequency

Shannon information indices have been widely employed in ecology but largely overlooked in genetics. Sherwin et al. (2006) assessed the performance, power and theoretical expectation of Shannon indices for estimating genetic diversity. They concluded that the Shannon information framework offers a powerful method of quantifying genetic diversity across multiple scales. Here we estimated Shannon's *Mutual Information* index, ${}^{S}H_{UA}$, between the sexes within sampling locations, and between sampling locations, for pairwise comparisons. A useful property of this index is that it can be directly transformed to the log-likelihood contingency test statistic G, providing a convenient way to test for differences in haplotype frequencies between the sexes. However, to avoid the risk of high type I error rates that have been reported under some conditions for G tests when employing the chi-square distribution to assess significance (see Ryman & Palm 2006), we test for statistical significance via random permutation (999 permutations).

Influence of small and uneven sizes

Sampling humpback whales along their migration can be strongly male biased, resulting from a potential excess of males heading north to the breeding grounds (Brown et al. 1995). Because an excess of males characterised much of our data, we were interested in assessing whether small sample sizes in some cases (from females) could bias our estimates of genetic differentiation. To test this, we re-ran the pairwise AMOVA analysis with the following modification. For each pairwise comparison of interest, we determined the smallest sample size for that pair and set that sample size for both groups. We then performed 1000 runs, each run randomly drawing without replacement from the available set of samples, up to the set sample size. We then computed the average F_{ST} and Φ_{ST} about these values across the 1000 runs. In this analysis, all individuals from the smallest group will be drawn from each run, with a different random subset of the samples drawn from the larger group.

RESULTS

I. Between sampling locations

The hierarchical AMOVA analysis found significant structure between the three sampling locations at the haplotype level (F_{RT} = 0.007, P = 0.001) but not at the nucleotide level (Table 2a). In pairwise comparisons, significant differentiation was detected between all sampling locations at the haplotype level but only between Stradbroke Island (STR) v. south-eastern Australia (SEA) and STR v. Evans Head (EVH) at the nucleotide level (Table 3a).

Pairwise haplotype sharing analyses found no significant differences between sampling locations whereas pairwise haplotype frequency analyses only found significant differences between the STR and SEA samples (${}^{S}H_{UA}$ = 0.191, P = 0.001; Table 3a).

II. Between sampling locations by migration direction

In northbound whales, the hierarchical AMOVA analysis at the haplotype level revealed significant structure between sampling locations overall (F_{RT} = 0.011, P = 0.030; Table

2b), with significance also detected between all sampling location pairwise comparisons (Table 3a). At the nucleotide level, the hierarchical AMOVA found no significant differentiation in northbound whales between sampling locations but in pairwise comparisons, significant differentiation was detected between STR v. SEA and STR v. EVH (Tables 2b and 3a).

In northbound whales, pairwise haplotype sharing analyses found no significant differences between sampling locations but only after sequential Bonferroni correction for STR v. SEA (P = 0.026; Table 3a). Haplotype frequency analysis found significant differences for northbound STR v. SEA (${}^{S}H_{UA}$ = 0.330, P = 0.008) and STR v. EVH whales (${}^{S}H_{UA}$ = 0.256, P = 0.027).

For southbound whales we found no evidence for significant differentiation between sampling locations from the hierarchical AMOVA (haplotype and nucleotide) or for all pairwise (STR v. SEA only) analyses (Tables 2c and 3a).

III. Between sexes within sampling locations

The hierarchical AMOVA analysis detected significant structure between sexes within sampling locations at the nucleotide level (Φ_{PR} = 0.026, P = 0.003) but not at the haplotype level (Table 2a)

Pairwise AMOVA comparisons only revealed significant differentiation between STR males and females at the nucleotide (Φ_{ST} = 0.059, P = 0.002; as reported by Valsecchi et al. 2010) but not the haplotype level (Table 3b). The haplotype sharing and haplotype frequency analyses found no significant differentiation between males and females at STR.

For EVH and SEA we found no evidence for significant differentiation between the sexes within these sampling locations from pairwise AMOVA (haplotype and nucleotide), haplotype frequency and haplotype sharing analyses (Tables 3b).

IV. Between sexes within sampling locations by migration direction

The hierarchical AMOVA analysis detected significant differentiation between sexes within sampling locations for northbound individuals at the nucleotide (Φ_{PR} = 0.049, P = 0.008) but not at the haplotype level (Table 2b). There was no significant differentiation between sexes for southbound whales at the haplotype and nucleotide levels (Table 2c).

As previously reported by Valsecchi et al. (2010), a pairwise AMOVA of northbound STR males and females revealed significant differentiation at the nucleotide level (Φ_{ST} = 0.114, P = 0.001; Table 3b). However, we found no significant differentiation at the haplotype level (F_{ST} = 0.026, P = 0.037) after correction for multiple comparisons (Table 3b).

In northbound whales at EVH and SEA we found no evidence for significant differentiation between the sexes within these sampling locations from pairwise AMOVA (haplotype and nucleotide), haplotype frequency and haplotype sharing analyses (Tables 3b).

Likewise, in southbound whales within the two sampling locations (STR and SEA), we also found no evidence for significant differentiation between the sexes from all pairwise analyses (Tables 3b).

Influence of small and uneven sample sizes

When the available data were partitioned by location, sex and migratory direction some sample sizes were small (min = 8; SEA northbound females). Simulation analyses that reduced the sample sizes in pairwise analyses to be equally small found the estimates of genetic differentiation changed little even for pairwise comparisons that included the smaller sample size (e.g. F_{ST} from -0.013 to -0.014 for SEA northbound males v. females; see Table 3, negatives not shown).

DISCUSSION

Our dataset of 315 samples gave us the opportunity to reassess the evidence for sex specific migration in humpback whales sampled at three locations along the eastern Australian migratory corridor. In our initial analysis of the mitochondrial sequence data, we were surprised to find significant structure among all sampling locations at the haplotype level, irrespective of sex or migratory direction, and for whales migrating north towards the breeding grounds. However, this result was only replicated at the nucleotide level for pairwise comparisons that included the STR sampling location. When samples were partitioned by sex and sampling location, we recovered the same intriguing finding as Valsecchi et al. (2010) of significant structure between STR males and females at the nucleotide level, both overall and for northbound individuals, but this was not repeated in the samples off EVH and SEA. Across all three eastern Australian sampling locations, regardless of migratory direction, we found no significant differences in the patterns of haplotype differentiation, haplotype sharing and haplotype frequency between the sexes. Overall, our results raise doubts about whether eastern Australian males and females are utilizing different migratory routes, as well as highlight other potential anomalies; both of which warrant closer scrutiny.

Those results that revealed significant differentiation either between sampling locations, or between sexes within a sampling location, almost exclusively involved the STR sample. To investigate this further we partitioned all haplotypes into the four primary clades (AE, CD, IJ and SH) described by Baker et al. (1993, 1998), and also employed by Valsecchi et al. (2010) to describe their data. When partitioned by sex and migration direction, northbound STR males have a near even representation of the CD (45.2%) and IJ clades (50.0%) whereas northbound females were predominantly of the CD clade (CD: 81%; IJ: 6.2%; see Table 4). The near even distribution of the CD and IJ clade was also found in southbound males (CD: 53%; IJ: 42.8%). For all the remaining samples, regardless of sex, migratory direction or sampling location, the distribution of the CD and IJ clades were uneven and remarkably similar, with approximately 70% CD (64-88%) and 25% IJ (12-28%; see Table 4). The STR result is also surprising as the frequency of the CD clade in other western South Pacific populations, including eastern Australia, has been previously reported to range between 60% and 85% (Palsbøll et al. 1995, Larsen et al. 1996, Baker et al. 1998, Olavarría et al. 2007).

To explain the differentiation between northbound males and females at STR, driven by the relatively high proportion of divergent haplotypes, Valsecchi et al. (2010) proposed and assessed several migration hypotheses. The one they considered most plausible suggests males within the eastern Australian breeding population consistently migrate north elsewhere in the western South Pacific but then migrate south off eastern Australia; females simply migrate north and south along eastern Australia. These authors argue that males may use such different migratory routes to females in order to 'maximise reproductive opportunities'. This implies males that ultimately return to the eastern Australian natal population do so after visiting females in other western Pacific breeding grounds. If this scenario is correct, we would expect to see substantial movement between the western South Pacific and eastern Australia.

The evidence for such movement is sparse. In a six year photo-identification study in the western South Pacific only four individuals were common to both the Oceania and eastern Australia catalogues (*N* = 672 and 1248 individuals respectively; Garrigue et al. 2011). In contrast, 20 individuals were resighted between breeding grounds in Oceania and 78 were resighted in the same breeding ground between seasons. Although a recent study factoring in the population size difference between Oceania (Constantine et al. 2012) and eastern Australia (Noad et al. 2011) found the movement rates among Oceania breeding grounds were no higher than between Oceania and eastern Australia, movement between all breeding grounds is still considered low (Jackson et al. 2012). Similarly, two individuals were resighted within-season between the Cook Strait (New Zealand) and 300kms north of Stradbroke Island suggesting a migratory link with eastern Australia (Franklin et al. In Press), but perhaps not the substantial movement required to produce the scenario proposed by Valsecchi et al. (2010). Genetic evidence also suggests that male migration behaviour is not facilitating broad scale gene flow through the western South Pacific. If there is substantial male movement within this region combined with female philopatry, as hypothesized by Valsecchi et al. (2010), we would expect significant spatial structure at mtDNA but not at nuclear markers. However, in a previous chapter of this thesis we found low but significant genetic differentiation between neighbouring breeding grounds using both mitochondrial and genomic markers (see Chapter 2; Table 2 and 3). Valsecchi et al. (2010) also imply that males migrating north off eastern Australia most likely belong to the New Caledonian breeding population. Yet in a simple examination of haplotype sharing, northbound STR males shared more haplotypes with males sampled elsewhere in eastern Australia than males sampled in the western South Pacific (see Table S3), with similar measures of F_{ST} for all comparisons (F_{ST} = 0.02, P < 0.05).

The lack of corroborating evidence for the Valsecchi hypothesis leads us to explore other possible explanations for the significant differentiation detected between males and females sampled off STR in 1992. When the available data were partitioned by sex and migration direction, the STR, and indeed the EVH and SEA samples sizes were relatively small (e.g. N = 16, STR females northbound). This raises the possibility that the lack of consistency in the in the measures of differentiation between males and females presented here are a product of statistical (Type I or II) error. Although female sample size was small in some analyses, females do not appear to be driving the Valsecchi et al. (2010) result. In the simple summary of haplotype frequencies (see Table 4), it is the STR males (arguably both north and southbound) that standout as anomalous for eastern Australia, yet both are represented by reasonable sizes (N = 42 and 49 respectively) and therefore less likely to be affected by sampling error. Likewise, STR males were found to be significantly differentiated from EVH males for both F_{ST} and Φ_{ST} (P < 0.001) despite the reasonable sample sizes.

If sample size is not an immediate concern, at least for males, then are there alternative explanations for the results presented here? One possible explanation is that some samples may not be representative of the population as a whole. This may occur if there is temporal genetic structure within the migration and sampling does not span the full migration period. For example, in a study of bowhead whales, Jorde et al. (2007) found elevated genetic differences between whales sampled a week apart on their migration past the coast of Barrow in Alaska. They could not provide a definitive explanation for the result but suspected it may be due to social structuring where related individuals migrate together or display similarity in their timing of migration. Alternatively they proposed that genetically differentiated whales may winter in different regions of the Bering Sea and thus may pass Barrow at different times or in separate pulses.

Migrating humpback whales show both age and sex segregation (Chittleborough 1965, Dawbin 1966, Craig et al. 2003) but are not known to migrate in kin-based pods (Valsecchi et al. 2002, Pomilla & Rosenbaum 2006). Rosenbaum et al. (2002) have also demonstrated that age-related genetic structure can occur due to interactions between reproductive success and environmental factors, and the impact of commercial whaling could accentuate age related structure. This suggests it may be possible to generate a sample that is not broadly representative of the migrating population if sampling is restricted to a short period within the migration. Valsecchi et al. (2002, 2010) state that samples were collected over a 69 day period during the peak of the northern and a 68 day period on the southern migration, although it is not clear what the distribution of sampling is within each of these periods. Sampling from EVH was also temporally restricted (16 day period) but extended in SEA (160 day period over two seasons). This limited period of sampling and temporal structure within a migration season may also explain why we detected weak but significant differentiation at the haplotype level between EVH and SEA (see Table 3), although this result was not repeated at the nucleotide level.

Another potential source of temporal genetic structure in the eastern Australian migratory corridor could be that more than one population uses this route either consistently or sporadically with temporal sex and age, or social structuring within and between the contributing populations. If this was the case however, then we would expect to see a greater number of resights between eastern Australia and New Caledonia than has been reported thus far (Garrigue et al. 2011).

In conclusion, a biological or statistical explanation for our findings and those of Valsecchi et al. (2010) remain uncertain and therefore we should remain circumspect about the possibility of sex specific migratory route choice along the eastern Australian migratory corridor. What is clear, is that the genetic differences observed between the sexes for humpback whales migrating past Stradbroke Island in 1992, using nucleotide level analyses, were not consistent within our eastern Australian sampling locations and across complimentary analyses. Resolving the potential complexities along the eastern Australian migratory corridor will involve improving the geographical and temporal coverage of humpback whale samples across the full extent of the migration, while also maintaining focus on possible mixing with neighbouring populations, age and social structure, and better understanding the influence of historical overexploitation and recovery on temporal structure. We would also recommend a multi-analysis approach, which includes non-genetic data such as satellite tagging to help avoid Type I errors, particularly when there is such weak differentiation characterising the breeding populations of the western South Pacific (Olavarría et al. 2007).

ACKNOWLEDGEMENTS

Thanks to the Australian Marine Mammal Centre at the Australian Antarctic Division for their financial support of the field and laboratory component of this project in collecting and processing the eastern Australian and Antarctic feeding ground samples. Thanks also to Valsecchi et al. (2010) for recommending a re-examination of sex-specific population dynamics in the western South Pacific.

LITERATURE CITED

Baker CS, Clapham PJ (2004) Modelling the past and future of whales and whaling. Trends Ecol Evol 19:365-371

Baker CS, Florez-Gonzalez L, Abernethy B, Rosenbaum HC, Slade RW, Capella J, Bannister JL (1998) Mitochondrial DNA variation and maternal gene flow among humpback whales of the Southern Hemisphere. Mar Mamm Sci 14:721-737

Baker CS, Medrano-Gonzalez L (2002) World-wide distribution and diversity of humpback whale mitochondrial DNA lineages. In: Pfeiffer CJ (ed) Molecular and Cell Biology of Marine Mammals. Krieger Publishing, Melbourne, FL, p 84-99

Baker CS, Perry A, Bannister JL, Weinrich MT, Abernethy RB, Calambokidis J, Lien J, Lambertsen RH, Ramirez JU, Vasquez O, Clapham PJ, Alling A, Obrien SJ, Palumbi SR (1993) Abundant Mitochondrial DNA Variation and World-Wide Population Structure in Humpback Whales. Proceedings of the National Academy of Sciences of the United States of America 90:8239-8243

Best P, Brandao, A. and Butterworth, D.S. (2001) Demographic parameters of southern right whales off South Africa. Journal of Cetacean Research and Management 2:161-169

Block BA, Jonsen ID, Jorgensen SJ, Winship AJ, Shaffer SA, Bograd SJ, Hazen EL, Foley DG, Breed GA, Harrison AL, Ganong JE, Swithenbank A, Castleton M, Dewar H, Mate BR, Shillinger GL, Schaefer KM, Benson SR, Weise MJ, Henry

RW, Costa DP (2011) Tracking apex marine predator movements in a dynamic ocean. Nature advance online publication

Brown M, Corkeron P (1995) Pod characteristics of migrating humpback whales (*Megaptera novaeangliae*) off the east Australia coast. Behaviour 132:163-179

Brown MR, Corkeron PJ, Hale PT, Schultz KW, Bryden MM (1995) Evidence for a sex-segregated migration in the humpback whale (*Megaptera novaeangliae*). Proceedings of the Royal Society of London Series B-Biological Sciences 259:229-234

Chittleborough RG (1965) Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). Australian Journal of Marine Freshwater Research 16:33-128

Clapham P (1996) The social and reproductive biology of humpback whales: an ecological perspective. Mammal Review 26:27-49

Clapham P, Robbins, J., Brown, M., Wade, P.R. and Findlay, K. (2001) A note on plausible rates on population growth in humpback whales. Appendix 5, Report of the Sub-Committee on the Comprehensive Assessment of Whale Stocks. Journal of Cetacean Research and Management 3:196-197

Clapham PJ, Baraff LS, Carlson CA, Christian MA, Mattila DK, Mayo CA, Murphy MA, Pittman S (1993) Seasonal occurrence and annual return of humpback whales, *Megaptera novaeangliae*, in the southern Gulf of Maine. Canadian Journal of Zoology 71:440-443

Clapham PJ, Mayo CA (1987) Reproduction and recruitment of individually identified humpback whales, *Megaptera novaeangliae*, observed in Massachusetts Bay, 1979-1985. Canadian Journal of Zoology 65:2853-2863

Clapham PJ, Seipt IE (1991) Resightings of independent fin whales, *Balaenoptera physalus*, on maternal summer ranges. Journal of Mammalogy 72:788-790

Constantine R, Jackson JA, Steel D, Baker CS, Brooks L, Burns D, Clapham P, Hauser N, Madon B, Mattila D, Oremus M, Poole M, Robbins J, Thompson K, Garrigue C (2012) Abundance of humpback whales in Oceania using photo-identification and microsatellite genotyping. Marine Ecology Progress Series 453:19

Cox GW (1985) The Evolution of Avian Migration Systems between Temperate and Tropical Regions of the New World. Am Nat 126:451-474

Craig AS, Herman LM (1997) Sex differences in site fidelity and migration of humpback whales (*Megaptera novaeangliae*) to the Hawaiian Islands. Canadian Journal of Zoology-Revue Canadienne De Zoologie 75:1923-1933

Craig AS, Herman LM, Gabriele CM, Pack AA (2003) Migratory timing of humpback whales (*Megaptera novaeangliae*) in the central North Pacific varies with age, sex and reproductive status. Behaviour 140:981-1001

Dawbin WH (1966) The seasonal migratory cycle of humpback whales. In: K.Norris (ed) Whales, Dolphins and Porpoises. University of California Press, Berkeley, California, p 145-170

Dingle H, Drake VA (2007) What is migration? Bioscience 57:113-121

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

Franklin W, Franklin T, Gibb N, Childerhouse S, Garrigue C, Constantine R, Brooks L, Burns D, Paton D, Poole M, Hauser N, Donoghue M, Russel K, Mattila D, Robbins J, Anderson M, Olavarria C, Jackson J, Noad MJ, Harrison P, Baverstock P, Leaper R, Baker CS, Clapham P (In Press) Photo-identification confirms that humpback whales (*Megaptera novaeangliae*) from eastern Australia migrate past New Zealand but indicates low levels of interchange with breeding grounds of Oceania. Journal of Cetacean Research and Management

Garrigue C, Forestell PH, Greaves J, Gill P, Naessig P, Patenaude NM, Baker CS (2000) Migratory movements of humpback whales (*Megaptera novaeangliae*) between New Caledonia, East Australia and New Zealand. Journal of Cetacean Research and Management 2:111-115

Garrigue C, Franklin T, Constantine R, Russel K, Burns D, Poole M, Paton D, Hauser N, Oremus M, Childerhouse S, Mattila D, Gibbs N, Franklin W, Robbins J, Clapham P, Baker CS (2011) First assessment of interchange of humpback whales between Oceania and the east coast of Australia. Journal of Cetacean Research and Management 3:269-274

Gronroos J, Green M, Alerstam T (2012) To fly or not to fly depending on winds: shorebird migration in different seasonal wind regimes. Animal Behaviour 83:1449-1457

Hudson RR, Boos DD, Kaplan N (1992) A statistical test for detecting geographic subdivisions. Molecular Biology and Evolution 9:138-151

Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss *Buchloe dactyloides* (Nutt) Engelm. Theor Appl Genet 86:927-934

IWC (2001) Report of the workshop on the comprehensive assessment of right whales: a worldwide comparison. Journal of Cetacean Research and Management 2:1-61

Jackson J, Anderson M, Steel D, Brooks L, Baverstock P, Burns D, Clapham P, Constantine R, Franklin W, Franklin T, Garrigue C, Hauser N, Paton D, Poole M, Baker CS (2012) Multistate measurements of genotype interchange between East Australia and Oceania (IWC breeding sub-stocks E1, E2, E3 and F) between 1999 and 2004, Paper SC/64/SH22 presented at the IWC Scientific Committee, Panama City, Panama

Jackson JA, Patenaude NJ, Carroll EL, Baker CS (2008) How few whales were there after whaling? Inference from contemporary mtDNA diversity. Mol Ecol 17:236-251

Jorde PE, Schweder T, Bickham JW, Givens GH, Suydam R, Hunter D, Stenseth NC (2007) Detecting genetic structure in migrating bowhead whales off the coast of Barrow, Alaska. Mol Ecol 16:1993-2004

Larsen AH, Sigurjonsson J, Oien N, Vikingsson G, Palsboll P (1996) Population genetic analysis of nuclear and mitochondrial loci in skin biopsies collected from central and northeastern North Atlantic humpback whales (*Megaptera novaeangliae*): population identity and migratory destinations. Proceedings of the Royal Society of London - Series B: Biological Sciences 263:1611-1618

Noad MJ, Cato DH, Paton D (2005) Absolute and relative abundance estimates of Australian east coast humpback whales (*Megaptera novaeangliae*), Paper SC/57/SH12 presented the IWC Scientific Committee, 2005 (unpublished).

Noad MJ, Dunlop RA, Paton D, Cato DH (2011) Absolute and relative abundance estimates of Australian east coast humpback whales (*Megaptera novaeangliae*). Journal of Cetacean Research and Management (special issue 3):243-252

Olavarría C, Baker CS, Garrigue C, Poole M, Hauser N, Caballero S, Flórez-González L, Brasseur M, Bannister J, Capella J, Clapham P, Dodemont R, Donoghue M, Jenner C, Jenner M-N, Moro D, Oremus M, Paton D, Rosenbaum H, Russell K (2007) Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. Marine Ecology Progress Series 330:257-268

Palsbøll PJ, Clapham PJ, Mattila DK, Larsen F, Sears R, Siegismund HR, Sigurjonsson J, Vasquez O, Arctander P (1995) Distribution of mtdna haplotypes in North Atlantic humpback whales - the influence of behaviour on population structure. Marine Ecology-Progress Series 116:1-10

Paterson R, Paterson P, Cato DH (1994) The status of humpback whales *(Megaptera novaeangliae)* in East Australia thirty years after whaling. Biological Conservation 70:135-142

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288-295

Peakall R, Smouse PE (2012) GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics 28:2537-2539

Peakall R, Smouse PE, Huff DR (1995) Evolutionary implications of allozyme and RAPD Variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. Mol Ecol 4:135-147

Pomilla C, Rosenbaum HC (2006) Estimates of relatedness in groups of humpback whales (*Megaptera novaeangliae*) on two wintering grounds of the Southern Hemisphere. Mol Ecol 15:2541-2555

Purcell J, Brodin A (2007) Factors influencing route choice by avian migrants: A dynamic programming model of Pacific brant migration. J Theor Biol 249:804-816

Rosenbaum HC, Weinrich MT, Stoleson SA, Gibbs JP, Baker CS, DeSalle R (2002) The effect of differential reproductive success on population genetic structure: Correlations of life history with matrilines in humpback whales of the Gulf of Maine. J Hered 93:389-399

Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Mol Ecol Notes 6:600-602

Schmitt NT, Double MC, Baker CS, Steel D, Jenner KCS, Jenner M-NM, Paton D, Gales R, Jarman SN, Gales N, Marthick JR, Polanowski AM, Peakall R (In Press) Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations. Mar Mamm Sci

Schmitt NT, Double MC, Gales N, Childerhouse S, Polanowski AM, Steel D, Albertson-Gibb R, Olavarria C, Garrigue C, Poole M, Hauser N, Constantine R, Paton D, Jenner C, Jarman SN, Peakall R (Chapter 3) Mixed-stock analysis of humpback whales (*Megaptera novaeangliae*) on Antarctic feeding grounds.

Senapathi D, Nicoll MAC, Teplitsky C, Jones CG, Norris K (2011) Climate change and the risks associated with delayed breeding in a tropical wild bird population. Proceedings of the Royal Society B-Biological Sciences 278:3184-3190

Sherwin WB, Jabot F, Rush R, Rossetto M (2006) Measurement of biological information with applications from genes to landscapes. Mol Ecol 15:2857-2869

Smith JN, Grantham HS, Gales N, Double MC, Noad MJ, Paton D (2012) Identification of humpback whale breeding and calving habitat in the Great Barrier Reef. Marine Ecology-Progress Series 447:259-272

Stanley CQ, MacPherson M, Fraser KC, McKinnon EA, Stutchbury BJM (2012) Repeat Tracking of Individual Songbirds Reveals Consistent Migration Timing but Flexibility in Route. PLoS ONE 7:e40688

Takahata N, Palumbi SR (1985) Extranuclear differentiation and gene flow in the finite island model. Genetics 109:441-457

Valsecchi E, Corkeron PJ, Galli P, Sherwin W, Bertorelle G (2010) Genetic evidence for sex-specific migratory behaviour in western South-Pacific humpback whales. Marine Ecology Progress Series 398:275-286

Valsecchi E, Hale P, Corkeron P, Amos W (2002) Social structure in migrating humpback whales (*Megaptera novaeangliae*). Mol Ecol 11:507-518

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

Yosef R, Markovets M, Mitchell L, Tryjanowski P (2006) Body condition as a determinant for stopover in bee-eaters (*Merops apiaster*) on spring migration in the Arava Valley, southern Israel. J Arid Environ 64:401-411

TABLES

Table 1. Description of samples used in analyses from individual whales sequenced at the mtDNA control region, M = males, F = females. (Number of males in parentheses)

Sample location	**Sampling period**	**Sex**		**Migratory direction**	
		M	**F**	**North (M)**	**South (M)**
Stradbroke Island	1992 (38 days - 5th June to 21st Oct)	91	44	58 (42)	77 (49)
Evan's Head	2009 (16 days - 16th June to 2nd July)	46	11	57 (46)	0
south-eastern Australia *	2006-2008	75	48	44 (36)	79 (39)
Eden	*2008 (14 days - 24th June to 1st Nov)*	*44*	*13*	*43 (35)*	*14 (9)*
Tasmania	*2007 (25 days - 6th July to 30th Nov)*	*31*	*35*	*1 (1)*	*65 (30)*

* Samples from Eden and Tasmania were pooled for this study following Schmitt et al. (In Press)

Table 2. Summary of hierarchical AMOVA analysis at the haplotype and nucleotide level, with molecular variance partitioned among the eastern Australian sampling locations and sex, with and without the influence of migratory direction. Significant P-values ($P < 0.05$) are based on statistical testing of 999 random permutations (calculated using the specialised permute option in GenAlEx 6.5).

a) *Not including migratory direction*

	Haplotype level			Nucleotide level		
Source	Statistic	Value	*P*	Statistic	Value	*P*
Among sampling locations	F_{RT}	0.007	0.001	Φ_{RT}	0.007	0.096
Among sexes within sampling locations	F_{PR}	0.001	0.334	Φ_{PR}	0.026	0.003

b) *For northbound whales*

	Haplotype level			Nucleotide level		
Source	Statistic	Value	*P*	Statistic	Value	*P*
Among sampling locations	F_{RT}	0.011	0.030	Φ_{RT}	0.006	0.314
Among sexes within sampling locations	F_{PR}	0.008	0.192	Φ_{PR}	0.049	0.008

c) *For southbound whales*

	Haplotype level			Nucleotide level		
Source	Statistics	Value	*P*	Statistics	Value	*P*
Among sampling locations	F_{RT}	0.002	0.287	Φ_{RT}	-0.001	0.549
Among sexes within sampling locations	F_{PR}	-0.003	0.766	Φ_{PR}	0.013	0.087

Table 3. Pairwise comparisions of haplotype sharing, genetic differentiation at the haplotype (F_{ST}) and nucleotide level (Φ_{ST}) and haplotype frequency differences (Shannon's mutual information index, *sHua*) between: a) the three eastern Australian sampling locations and, b) male and female humpback whales within each sampling location, with and without the influence of migratory direction. We also include average F_{ST} and Φ_{ST} values based on sub-sampling with equal sample sizes (ES), with 95% confidence intervals in parentheses. The significance levels shown for genetic differentiation, haplotype sharing analysis and haplotype frequency are based on the outcomes of statistical tests by random permutation (999 permutations). Significant P-values (AMOVA, $P < 0.05$; 2-tailed non-parametric, $P < 0.025$) are shown in bold after correction for multiple comparison with the sequential Bonferroni test (Rice 1989).

STR = Stradbroke Island; EVH = Evan's Head; SEA = south-eastern Australia.

a)

	Haplotype Sharing			Genetic Differentiation						Haplotype frequencies	
Sampling location	No. of shared haplotypes/total	P (rand >= data)	P (rand <= data)	F_{ST}	F_{ST} (ES)	P	Φ_{ST}	Φ_{ST} (ES)	P	*sHua*	Permutation probability
STR vs EVH	22/47	0.400	0.782	**0.008**	**0.008 (0, 0.018)**	**0.027**	**0.039**	**0.041 (0.015, 0.073)**	**0.001**	0.170	0.117
STR vs SEA	27/55	0.150	0.946	**0.007**	**0.007 (0.005, 0.008)**	**0.002**	**0.019**	**0.019 (0.015, 0.024)**	**0.004**	**0.191**	**0.001**
EVH vs SEA	20/47	0.462	0.765	**0.009**	**0.009 (0.001, 0.019)**	**0.016**	0.006	0.006 (0, 0.022)	0.167	0.167	0.300
Northbound whales											
STR vs EVH	18/33	0.308	0.867	**0.018**	**0.018 (0.018, 0.018)**	**0.009**	**0.057**	**0.057 (0.057, 0.057)**	**0.001**	0.256	0.027
STR vs SEA	13/36	0.026	0.995	**0.014**	**0.014 (0.007, 0.022)**	**0.016**	**0.035**	**0.035 (0.018, 0.056)**	**0.023**	**0.330**	**0.008**
EVH vs SEA	17/35	0.835	0.383	**0.013**	**0.013 (0.003, 0.024)**	**0.022**	0.014	0.014 (0, 0.034)	0.101	0.238	0.270
Southbound whales											
STR vs SEA	25/48	0.871	0.325	0	0 (0,0)	0.453	0.006	0.006 (0.001, 0.012)	0.190	0.124	0.627

b)

	Haplotype Sharing			Genetic Differentiation						Haplotype frequencies	
Sampling location and sex	No. of shared haplotypes/total	P (rand >= data)	P (rand <= data)	F_{ST}	F_{ST} (ES)	P	Φ_{ST}	Φ_{ST} (ES)	P	*sHua*	Permutation probability
STR males vs STR females	17/42	0.226	0.914	0.006	0.006 (0, 0.016)	0.087	**0.059**	**0.061 (0.025, 0.109)**	**0.002**	0.217	0.162
EVH males vs EVH females	8/27	0.962	0.147	0	0 (0, 0.053)	0.420	0	0 (0, 0.058)	0.390	0.224	0.686
SEA males vs SEA females	17/40	0.268	0.873	0	0 (0, 0.004)	0.460	0	0 (0, 0.016)	0.390	0.198	0.540
Northbound whales											
STR males vs STR females	8/24	0.398	0.807	0.026	0.026 (0, 0.058)	0.037	**0.114**	**0.111 (0.029, 0.218)**	**0.001**	0.318	0.082
SEA males vs SEA females	4/25	0.558	0.766	0	0 (0, 0.036)	0.457	0	0 (0, 0.116)	0.396	0.266	0.611
EVH males vs EVH females	8/27	0.967	0.137	0	0 (0, 0.052)	0.442	0	0 (0, 0.057)	0.389	0.224	0.677
Southbound whales											
STR males vs STR females	12/38	0.600	0.621	0	0 (0, 0.002)	0.457	0.016	0.015 (0, 0.063)	0.215	0.286	0.726
SEA males vs SEA females	14/35	0.664	0.561	0	0 (0, 0)	0.456	0.011	0.011 (0.007, 0.017)	0.157	0.281	0.570

Note: all negative F_{ST} and Φ_{ST} values have been converted to 0

Table 4. The number of mtDNA control region haplotypes represented for humpback whales sampled off Stradbroke Island (Valsecchi et al. 2010), Evan's Head, and south eastern Australia (Schmitt et al. In review) for all samples and subclasses within each dataset. The columns on the right-hand side shows the number of individuals in each of four clades (AE, CD, IJ and SH) described in previous studies (Baker et al. 1993; 1998; Olavarria et al. 2007), and the % of individuals in the CD and IJ clades. *N* = number of samples.

Sampling location	**Sex**	**Migratory direction**	***N***	**No. of haplotypes**	**AE/CD/IJ/SH**	**%CD**	**%IJ**
Stradbroke Island	M	north	42	23	0/19/21/2	45.2	50.0
	F	north	16	10	0/13/1/2	81.3	6.3
	M	south	49	30	0/26/21/2	53.1	42.9
	F	south	28	19	0/20/7/1	71.4	25.0
Evan's Head	M	north	46	25	0/36/8/2	78.3	17.4
	F	north	11	10	0/8/3/0	72.7	27.3
south-eastern Australia	M	north	36	22	0/25/10/1	69.4	27.8
	F	north	8	7	0/7/1/0	87.5	12.5
	M	south	39	26	0/25/11/3	64.1	28.2
	F	south	40	23	0/31/9/0	77.5	22.5

Table S1. Summary of hierarchical AMOVA analysis at the haplotype and nucleotide level, with estimates of variance components made among all eastern Australian sampling locations (*RT*), among sexes within sampling locations (*PR*), and among sexes (*PT*). Significant P-values ($P < 0.05$) are based on statistical testing of 999 random permutations (calculated using the specialised permute option in GenAlEx 6.5).

Haplotype level

Source	*df*	SS	MS	Est. Var.	%
Among sampling locations	2	1.691	0.845	0.003	1%
Among sexes within sampling locations	3	1.505	0.502	0.000	0%
Among sexes	309	148.794	0.482	0.482	99%
Total	314	151.990		0.485	100%

Nucleotide level

Source	*df*	SS	MS	Est. Var.	%
Among sampling locations	2	29.329	14.664	0.033	1%
Among sexes within sampling locations	3	29.991	9.997	0.121	3%
Among sexes	304	1392.726	4.581	4.581	97%
Total	309	1452.045		4.736	100%

Statistic	Value	*P*	Statistic	Value	*P*
F_{RT}	0.007	0.001	Φ_{RT}	0.007	0.096
F_{PR}	0.001	0.334	Φ_{PR}	0.026	0.003
F_{PT}	0.008	0.002	Φ_{PT}	0.033	0.001
F'_{RT}	0.189		Φ'_{RT}	0.008	
F'_{PR}	0.026		Φ'_{PR}	0.031	

*Due to the limitations of F- and Φ-statistics for analysing data from highly variable loci (Hedrick 2005; Meirmans & Hedrick 2010) we also present the standardised F'- and Φ'-statistics. These are calculated by standardising the normal F- and Φ-statistics by the maximum value it can obtain given the observed within population diversity.

Table S2a. Summary of hierarchical AMOVA analysis at the haplotype and nucleotide level, with estimates of variance components made among all eastern Australian sampling locations (RT), among sexes within sampling locations (PR) and among sexes (PT) for northbound humpback whales. Significant P-values ($P < 0.05$) are based on statistical testing of 999 random permutations (calculated using the specialised permute option in GenAlEx 6.5).

Haplotype level

Source	*df*	SS	MS	Est. Var.	%
Among sampling locations	2	1.737	0.868	0.005	1%
Among sexes within sampling locations	3	1.623	0.541	0.004	1%
Among sexes	152	72.425	0.476	0.476	98%
Total	157	75.785		0.485	100%

Nucleotide level

Source	*df*	SS	MS	Est. Var.	%
Among sampling locations	2	28.229	14.115	0.029	1%
Among sexes within sampling locations	3	26.129	8.710	0.234	5%
Among sexes	152	685.863	4.512	4.512	94%
Total	157	740.222		4.775	100%

Statistic	Value	*P*	Statistic	Value	*P*
F_{RT}	0.011	0.030	Φ_{RT}	0.006	0.314
F_{PR}	0.008	0.192	Φ_{PR}	0.049	0.008
F_{PT}	0.018	0.001	Φ_{PT}	0.055	0.001
F'_{RT}	0.267		Φ'_{RT}	0.007	
F'_{PR}	0.167		Φ'_{PR}	0.059	

*Due to the limitations of F- and Φ-statistics for analysing data from highly variable loci (Hedrick 2005; Meirmans & Hedrick 2010) we also present the standardised F'- and Φ'-statistics. These are calculated by standardising the normal F- and Φ-statistics by the maximum value it can obtain given the observed within population diversity.

Table S2b. Summary of hierarchical AMOVA analysis at the haplotype and nucleotide level, with estimates of variance components made among all eastern Australian sampling locations (*RT*), among sexes within sampling locations (*PR*) and among sexes (*PT*) for southbound humpback whales. Significant P-values ($P < 0.05$) are based on statistical testing of 999 random permutations (calculated using the specialised permute option in GenAlEx 6.5).

Haplotype level

Source	*df*	SS	MS	Est. Var.	%
Among sampling locations	1	0.487	0.487	0.001	0%
Among sexes within sampling locations	2	0.849	0.424	0.000	0%
Among sexes	153	74.461	0.487	0.487	100%
Total	156	75.796		0.488	100%

Nucleotide level

Source	*df*	SS	MS	Est. Var.	%
Among sampling locations	1	6.684	6.684	0.000	0%
Among sexes within sampling locations	2	13.985	6.993	0.064	1%
Among sexes	148	688.896	4.655	4.655	99%
Total	151	709.566		4.718	100%

Statistics	Value	*P*	Statistics	Value	*P*
F_{RT}	0.002	0.287	Φ_{RT}	-0.001	0.549
F_{PR}	-0.003	0.766	Φ_{PR}	0.013	0.087
F_{PT}	-0.002	0.649	Φ_{PT}	0.012	0.076
F'_{RT}	0.059		Φ'_{RT}	-0.002	
F'_{PR}	-0.126		Φ'_{PR}	0.016	

*Due to the limitations of F- and Φ-statistics for analysing data from highly variable loci (Hedrick 2005; Meirmans & Hedrick 2010) we also present the standardised F'- and Φ'-statistics. These are calculated by standardising the normal F- and Φ-statistics by the maximum value it can obtain given the observed within population diversity.

Table S3. The number of shared haplotypes as a proportion of the total between northbound Stradbroke Island males (STR), northbound males from south-eastern Australia (SEA) and Evan's Head (EVH), and males from neighbouring breeding grounds.
WA = western Australia; NC = New Caledonia; TG = Tonga

Haplotype sharing between northbound STR males and:

Sampling location and sex	No. of shared haplotypes/total	No. of shared haplotypes/total (proportion)
SEA males (northbound)	11/33	0.33
EVH males (northbound)	15/32	0.47
WA males	11/55	0.20
NC males	17/63	0.27
TG males	16/55	0.29

Chapter 4: Development and evaluation of single nucleotide polymorphism (SNP) markers for population structure analysis in the humpback whale (*Megaptera novaeangliae*)

Natalie T. Schmitt[1,2], Andrea M. Polanowski[1,4], Michael C. Double[1], Scott Baker[3], Debbie Steel[3], Rod Peakall[2], Simon N. Jarman[4]

[1] Australian Marine Mammal Centre, Australian Antarctic Division, Kingston, Tasmania 7050, Australia

[2] Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra ACT 0200, Australia

[3] Marine Mammal Institute and the Department of Fisheries and Wildlife, Oregon State University, 2030 SE Marine Science Dr, Newport Oregon 97365 USA

[4] Australian Antarctic Division, Kingston, Tasmania 7050, Australia

ABSTRACT

Studies of cetacean population genetics would benefit from the use of SNPs as scoring is more reproducible than for microsatellites, making them suitable for long-term, collaborative studies of globally distributed species. Here we report the development of SNPs in the humpback whale and the assessment of their statistical power for population genetic structure analysis. *Taqman*® assays for 17 SNPs and ten microsatellite loci were used to genotype samples from western Australia (*n*=22), eastern Australia (*n*=23) and California (*n*=22). POWSIM was used to simulate the statistical power of both markers to detect population structure for different values of F_{ST}, based on empirical estimates of allele frequencies. The simulations suggest adequate power to detect F_{ST} >0.03 for both markers, given the sample sizes of this study. For the detection of structure (F_{ST} = 0.005) that is typical among some humpback populations, sample sizes >150 would be required for the SNPs, or *N*>75 if 50 loci are used, compared to *N*>50 for the microsatellites. Consistent with the simulation, an AMOVA found significant differentiation between California and each of the two Australian regions at the microsatellites (F_{ST} = 0.029 and 0.036), but only between eastern Australia and California using the SNPs (F_{ST} = 0.034).

Keywords: humpback whale; population genetics; SNPs; microsatellites; POWSIM

INTRODUCTION

Studies of cetacean population genetics are often limited by the number and geographic coverage of available samples because of the costs and logistic challenges of collecting biopsy samples from individuals that are wide-ranging and at times elusive. For this reason, studies are often carried out over many years and across multiple laboratories. To date, nuclear microsatellite loci and mitochondrial DNA (mtDNA) sequences have been the genetic markers of choice for humpback whale (*Megaptera novaeangliae*) population genetic studies. However, microsatellites can be problematic in long-term population studies that require aquisition of data over time, and from multiple labs, due to the problem of inconsistencies in allele amplification and size calling (LaHood *et al.* 2002, Vignal *et al.* 2002, Hoffman *et al.* 2006). This problem has been overcome for human forensic analysis by the use of coordinated international standards involving the strategic selection of ‘reliable’ loci (Parentage Testing Committee 2003, Butler 2005). Such standardization is not easily implemented for the studies of cetaceans in part at least because of restrictions on international transfer of samples and DNA under the Convention on International Trade in Endangered Species (CITES), inconsistencies in technology between laboratories and the presence of markers that are often difficult to genotype. Programs such as Allelogram (Morin *et al.* 2009a) attempt to resolve these concerns by using controls to normalize and bin alleles from multiple sources. Mitochondrial markers are more prevalent among humpback whale population studies as they are more comparable, with nine studies using data derived from multiple laboratories compared to three for microsatellites from 2004 to 2011. However, as a single maternally inherited locus, it complements but cannot substitute for variable nuclear markers (Gagneux *et al.* 1997, Harpending *et al.* 1998, Hare 2001).

Single nucleotide polymorphisms (SNPs) have the potential to be genotyped with greater reliability than microsatellites as they are less technology dependent, and are especially useful when datasets need to be combined across time and between laboratories (Vignal *et al.* 2002, Brumfield *et al.* 2003, Morin *et al.* 2004). This is because SNP genotypes are detected as sequence differences rather than estimated allele sizes. Although SNPs generally have only two alleles compared with the multiallelic and often hypervariable microsatellites, their high abundance across the genome, and their ability to statistically outperform microsatellites with sufficient loci and appropriate sample sizes, make them a useful marker in population structure analysis (Liu *et al.* 2005, Lao *et al.* 2006, Paschou *et al.* 2007, Ryynanen *et al.* 2007, Sacks and Louie 2008, Freamo *et al.* 2011). The mutation rates and complex mutation models associated with microsatellite markers can also be difficult to predict, frequently leading to homoplastic alleles which may be identical in size but not descent (Macaubas *et al.* 1997). The presence of size homoplasy in a microsatellite data set is likely to dampen the signal of population structure, especially for large-scale population comparisons or comparisons between species (Haasl and Payseur 2011). Although homoplasy can also be an issue with bi-allelic markers, the frequency is likely to be reduced due to their lower mutation rate and simple, well defined mutation model. Divergence-based statistics like F_{ST} and D_{EST} are therefore easier to interpret for SNPs without concern for the impacts of within-group polymorphism or incorrect mutation model (Meirmans and Hedrick 2011).

Nuclear intron sequences have been used previously in the study of cetacean population structure (e.g. Palumbi and Baker 1994) and more recently as a source of variation for SNP discovery in the finless porpoise (Li *et al.* 2009), sperm whales (Morin *et al.*

2007a), and in bowhead whales for population structure analysis (Morin *et al.* 2010). Intronic SNPs have also had broad taxonomic utility across a diversity of more distantly related taxa such as the red fox and Pacific salmon (Smith *et al.* 2004, Smith *et al.* 2005, Pariset *et al.* 2006, Sacks and Louie 2008, Sevane *et al.* 2010). In this study, we report for the first time the development and testing of SNP markers for humpback whales. Our specific objectives were; 1) to develop a suite of informative SNP markers from intron sequences; 2) estimate by computer simulation the statistical power of a combined panel of intronic and exonic SNPs to detect population genetic structure compared with a suite of microsatellite markers, using an approach utilized by Morin *et al.* (2009b); 3) compare the results of the simulation with an analysis of molecular variance between closely related and distant humpback populations.

MATERIALS AND METHODS

Sample collection and DNA extraction

The sample set consisted of 67 humpback whale skin biopsy samples collected from neighbouring regions eastern Australia (Eden, NSW; 37° 3' S, 150° E; *N* = 18 and Evan's Head, NSW; 29° 7' S, 153° 27' E; *N* = 5) and western Australia (Exmouth; 21° 55' S, 114° 01' E; *N* = 22), and a more distant region off the coast of central California in the North Pacific (37° 48' N, 122° 24' W; *N* = 22) (Baker *et al.* 1998, Schmitt *et al.* In Press). Based on previous microsatellite analysis the dataset was known *a priori* to include no duplicate samples. However, the western Australian samples strategically included three mother/calf pairs to assist in the detection of genotyping errors. Total cellular DNA for the Australian samples was extracted from skin tissue using an automated Promega Maxwell ® 16 System. The central California samples were provided by Oregon State University as whole genome amplified DNA (WGA), amplified using Templiphi V2 from GE Biosciences.

Intronic SNPs

For intronic SNP discovery, 'exon priming intron crossing' (EPIC, see Lessa 1992, Slade *et al.* 1993, Palumbi and Baker 1994) primer sets were designed from humpback whale transcriptome data aligned to chromosomes 1, 2 and 3 of the *Bos taurus* genome using the software Bowtie (Langmead *et al.* 2009). The transcriptome data was assembled from 17,000,000 Illumina 65 bp unpaired reads from a pool of cDNA created from the epidermal tissue of six adult whales. Fragments for exon primer design were selected at random from each of the three Bovine chromosomes. Potential intronic regions of at least 600bps in length were chosen to maximise the sequencing read of a PCR product with long enough exonic sequences flanking both sides of the desired intron for PCR primer design. Primer pairs were BLASTed in GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to minimize multiple targets within the cow genome and therefore, hopefully the humpback genome.

In total, 41 epic primer pairs were designed. These were then tested for sequencing suitability using total cellular DNA and a polymerase chain reaction (PCR) gradient thermocycle with annealing temperatures ranging from 59˚C to 66˚C. Loci were classified (0-3) according to whether they produced (0) no product; (1) a single band; (2) two bands; or (3) three or more bands via agarose gel electrophoresis. Loci producing a single band were considered 'suitable' for sequencing and loci producing a double band could be optimized potentially to one band by gel purification (Freeze N'

Squeeze DNA Gel Extraction®Bio-Rad). Amplifications were conducted in a final volume of 10µl at the following concentrations: 2.5mM $MgCl_2$, 200µM dNTP, 0.2mM each primer, 0.25 U Taq (New England BioLabs ®Inc.), 1 X PCR reaction buffer (10mM Tris-HCl, 50mM KCl, 1.5mM $MgCl_2$), 10 X BSA and 1µl DNA (approximately 5-10ng). Temperature profiles consisted of an initial denaturing period of 2 min at 94˚C, followed by 35 cycles of denaturation at 94˚C for 30 s, annealing at 59-66˚C for 40 s, and extension at 72˚C for 40 s. A final extension period for 10 min at 72˚C was also included.

Successfully optimized loci were used to amplify a subset of eight to 24 individuals (average 14) representing the humpback whales that migrate along the east and west coast of Australia, to identify variable sites (Table 1). We chose samples that were representative of each Australian region in the discovery panel in an attempt to minimise ascertainment bias influencing subsequent analyses. These amplifications followed the above conditions with optimized annealing temperatures determined from the initial PCRs (Table 2). Unincorporated primers (and nucleotides) were removed from PCR amplicons using ExoSAP-IT®. Sequencing reactions were run on a standard thermocycler using either the forward or reverse primer and a Big Dye terminator sequencing kit (Applied Biosystems). Agencourt CleanSEQ was used for the post-sequencing removal of unincorporated primers. PCR products were sequenced on an ABI 3130 automated sequencer and then aligned within Sequencher®4.8 (Gene Codes Corp.). Chromatograms were visually inspected for sequencing errors. Sequences containing SNPs were then BLASTed against public database sequences in GenBank to assess identity of the target gene (Table 1) (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

SNPs were considered real (as opposed to sequencing artefacts) if they were detected in both sequence strands, or in one direction only but from either multiple individuals or a region of good quality sequence (i.e. single product with reliable signal). We considered a site to be polymorphic only if the height of the secondary peak was at least 75% of the height of the primary peak. Loci that provided poor sequence (i.e. amplified multiple products, or produced a very low, unreliable signal resulting from interspersed repetitive regions) were discarded from further analyses. SNPs were selected for further analyses based on sufficient flanking sequence to allow design of PCR primers (>100bps), and when two SNPs were used from a single locus, evidence that the SNPs were not in phase (i.e. > two haplotypes present in the sample subset). To minimise ascertainment bias, SNPs were defined as all variable sites within a sequence and not just those with high heterozygosity or displaying all three genotypes (Brumfield *et al.* 2003).

TaqMan® PCR probes and 'custom Taqman SNP genotyping assays' were designed by Applied Biosystems and used to genotype all SNPs. An exonic SNP marker discovery project was also conducted in parallel to this study (Polanowski *et al.* 2011) and we were able to incorporate ten of these exonic SNPs to increase the total number of loci (see Table 3 for TaqMan primer sets). Our goal was to obtain ten intronic SNPs along with the ten exonic SNPs. SNPs were genotyped for all 67 humpback whale samples. Genomic DNA (gDNA) from eastern and western Australia was re-extracted using the CTAB extraction method (Winnepenninckx *et al.* 1993) to increase DNA concentration and quality and then standardised to 2.5ng/µl per sample as recommended by the kit protocol (Custom TaqMan® SNP Genotyping Assays Protocol, Applied Biosystems; 1-20ng/µl). Whole genome amplified (WGA) DNA from central California was validated by genotyping a suite of ten microsatellite loci, following the methods of Schmitt *et al.*

(In Press), and comparing the results to those from the original genomic DNA (genotyped at Oregon University). The WGA DNA was then also standardised to 2.5ng/µl. qPCR amplifications were conducted in a 10ul reaction volume containing: 0.25 µl of 40XTaqMan SNP Genotyping assay (consisting of forward and reverse primers at 36µM concentration, and two TaqMan MGB probes at 8µM concentration, designed by Applied Biosystems), 5µl of 2XTaqMan SNP Genotyping Master Mix (Applied Biosystems) and 4µl of 2.5ng/µl DNA template. A negative control was included by adding sterile deionised water instead of DNA template. Real-time PCR was performed in optical 96-well reaction plates on the Roche LightCycler®480 using the manufacturers instructions for TaqMan probes.

SNP Validation

Four steps were taken to determine whether our suite of SNP markers were robust for population structure analyses.

1. To estimate genotyping error rate, a subset of 20 humpback samples from Australia and five from central California were randomly selected then re-genotyped and scored by an independent person at all SNP loci.

2. Three mother/calf pairs were included in the western Australian dataset to ensure genotypes were consistent with Mendelian single-locus inheritance across all loci.

3. Tests for deviation from Hardy-Weinberg equilibrium at each locus and linkage disequilibrium between loci were performed using Arlequin 3.1 (Excoffier *et al.* 2005), and Haploview v4.2 (Barrett *et al.* 2005) respectively. Sequential

Bonferroni correction was applied to all multiple pairwise comparisons (Rice 1989).

4. To provide an independent estimate of population structure for the same set of samples, we also compared the SNP results with those from a panel of ten microsatellite loci from Schmitt *et al*. (In Press).

Data analysis

I. Genetic Diversity

The general characteristics of each of the 17 SNP loci such as minor allele frequencies, the observed and expected heterozygosity were determined for each region using GenAlEx 6.3 (Peakall and Smouse 2006).

II. Simulation of statistical power

The program POWSIM (Ryman and Palm 2006) was used to assess the statistical power of the 17 SNP loci to detect different levels of structure between two populations based on the allele frequencies of the Australian dataset. The performance of the 17 SNP loci was then compared to that of ten microsatellite markers. The program mimics sampling from the populations at predefined *expected* levels of divergence (measured as F_{ST}) using information on effective population size, number of samples, number of loci, and allele frequencies for any hypothetical degree of true differentiation. Allele frequency drift was simulated under a Wright-Fisher model without mutation or gene flow subsequent to divergence. The expected level of differentiation between the populations is determined by the number of generations that have diverged according to the formula:

F_{ST} = 1-(1-1/(2N_e))t,

where N_e = effective population size and t = number of generations. The significance of the tests is assessed with χ^2 as well as Fisher's exact tests. Weak but significant differentiation (F_{ST} of 0.0025 to 0.025) is a common characteristic of neighbouring humpback whale regions in the southern hemisphere (Schmitt *et al.* In Press). Given this finding, we simulated genetic drift to expected F_{ST} levels representing weak (0.001, 0.0025, 0.005, 0.01 and 0.02), and moderate (0.04, 0.05 and 0.06) differentiation for both marker types by setting the numbers of generations (t), with a sample size of 23 for each region (approximately consistent with sample sizes from this study), using an effective population size (N_e) of 1000. We also tested the effect of population sample size on the statistical power of each marker to detect population structure at an expected level of F_{ST} = 0.005. For the SNP markers we then examined the relationship between population sample size and the number of loci needed to detect an F_{ST} of 0.005, using average allele frequencies from the Australian dataset. The statistical α-error, or the probability of rejecting H_O when it is true, was estimated by omitting genetic drift (i.e. F_{ST} = 0) as per Ryman and Palm (2006).

III. Population differentiation

The extent of genetic differentiation among all three regions was assessed for both the SNP and microsatellite loci in an Analysis of Molecular Variance (AMOVA) as implemented in GenAlEx 6.3 (Peakall and Smouse 2006), with statistical testing by random permutation (999 permutations). The AMOVA analysis provided an estimate of F_{ST} (under the infinite allele model) for codominant data following Peakall *et al.* (1995) and Michalakis and Excoffier (1996).

Given a recent suggestion that F_{ST} may not offer the best estimation of population divergence when using highly variable markers like microsatellites (Hedrick 2005, Jost 2008, Meirmans and Hedrick 2011), Jost's D_{EST} was also calculated for both markers using a modified version of the R package DEMEtics V0.8.0 (Jueterbock *et al.* 2010), with statistical testing by bootstrapping with 1000 permutations. Compared with F_{ST}, D_{EST} partitions diversity based on the effective number of alleles rather than on the expected diversity to give an unbiased estimation of divergence (Jost 2008; but see Meirmans and Hedrick 2011, for alternative estimators).

RESULTS

Intron sequencing

Of the 41 primer pairs developed from cDNA aligned to Bovine chromosomes 1-3, 26 loci yielded 'usable' amplification, with all but two of these loci producing a single strong band as visualised by agarose gel electrophoresis. Of these, 21 loci produced good quality sequence for SNP detection. Of the loci containing SNPs, nine (82%) had somewhat similar matches (>74%) to known sequence regions in other mammals with two loci (18%) closely matching to regions of the Bovine genome (>88%) following a BLAST search (Table 1).

Intronic SNP detection

A total of 16 variable sites (Table 2), or potential SNPs, were found among eleven loci and 13771 bps of intronic sequence, or one SNP every 861 bps. However, polymorphism was not uniformly distributed across all loci, with three loci containing multiple variable sites. Most SNPs were identified using eight individual samples.

Primer sequences, annealing temperatures and approximate fragment length can be found in Table 2. Potential SNPs could be classified into one of three categories (i) The SNP was visible in both directions; (ii) the SNP appeared in only one direction, in high quality sequence, and the reverse sequence was not available; and (iii) the SNP was visible in only one direction but either (a) the reverse sequence contradicted it, or (b) the reverse direction was not available and sequence quality was not high. In summary, ten SNPs across nine intronic loci (two loci were linked) fell into categories 1 or 2 and were selected for genotyping, and six other possible SNPs were not verified by further sequencing or locus optimization. Among the ten SNPs identified positively, eight were transitions and two tranversions.

Assay development and SNP validation

We developed TaqMan® assays for ten intronic and ten exonic SNP loci from which there was sufficient high quality sequence for primer design (Table 3). Three intronic assays were removed from further analyses because they failed to amplify consistently or failed to produce interpretable genotypes, or they deviated significantly from Hardy-Weinberg equilibrium for the 45 Australian humpback samples (data not shown). The remaining 17 SNP assays were genotyped for all 67 samples representing three humpback whale regions. Of these 17 SNPs, only one pair shared the same sequence fragment (UHMK1 and UHMK2, 155 bps apart). The WGA DNA from the central Californian samples was shown to be reliable for genotyping by producing results consistent to the genomic DNA samples when genotyped at ten microsatellite loci. GenBank Accession numbers for the sequences generated from these 17 loci are: 410759353 – 410759359 (intronic), 410759319 – 410759325 (exonic), 410759331 – 410759333 (exonic).

The 17 SNP loci finally selected were in Hardy-Weinberg equilibrium for both Australian regions and pairwise comparisons between loci revealed no significant linkage disequilibrium (all values of $P > 0.05$) for all three regions after sequential Bonferroni correction (Table 4). The power to detect linkage between the two loci that share the same sequence fragment was limited (UHMK1 and UHMK2) due to UHMK1 being monomorphic for western Australia and displaying low variation for the other two regions. TSPAN3 was also found to be monomorphic for California. Two loci for central California were found to deviate significantly from Hardy-Weinberg equilibrium in this region (P2YS and ZRANB2). However, analyses were run with and without these loci and in each case showed similar results and *P*-values (data not shown). Repeat genotyping of 20 samples from Australian humpback whales and five from California by an independent person revealed no inconsistencies across 850 allele calls. The three mother/calf pairs from western Australia also showed genotypes consistent with Mendelian single-locus inheritance across all 17 loci. Following validation, we removed one from each pair of mother/calf samples, leaving 64 samples in total.

Data analyses

I. Genetic Diversity and simulation of statistical power

All SNPs were bi-allelic with minor allele frequency ranges and averages similar for all three humpback regions: EA from 0.02 to 0.48 with an average of 0.25; WA from 0 to 0.47 with an average of 0.24; Cal from 0 to 0.5 with an average of 0.25 (Table 4). Average expected heterozygosity across loci were also similar for all regions (EA: 0.32 ± 0.04; WA: 0.32 ± 0.04; Cal: 0.32 ± 0.04).

For the simulation results we chose to report the most powerful tests for each set of markers: the Fisher's exact test for the microsatellite loci and the Chi-squared contingency test for the SNP markers, as recommended by Ryman *et al*. (2006). The estimate of the probability of rejecting the null hypothesis (H_O; genetic homogeneity) when it is true was 0.040 for the 17 SNPs and 0.051 for the microsatellite markers, as expected for a reliable test when using the 0.05 limit for significance (Figure 1). The probability of obtaining a significant result when sampling from 23 individuals per region, at Ne = 1000, exceeded 80% at the expected F_{ST} values of >0.01 and >0.03, for the ten microsatellites and 17 SNPs respectively (Figure 1). Thus the simulations indicate that the ten microsatellite loci offer more power to detect lower levels of differentiation than the 17 SNPs.

Simulations of the effect of the number of genotyped individuals within each population on the statistical power to detect differentiation at an expected the level of F_{ST} = 0.005, for both the SNPs and microsatellite markers, are shown in Figure 2. To detect differentiation between populations with a >80% probability for an expected F_{ST} value as low as 0.005, sample sizes of >50 would be required for the ten microsatellite markers, while for the 17 SNPs >150 samples are needed to ensure reliable detection. Thus as expected, larger sample sizes are required to detect low levels of differentiation by 17 SNPs, compared with ten microsatellites. By increasing the number of independent SNP loci to 50, using a minor allele frequency of 0.25 in the simulation, >75 samples are required to accurately detect (>80% probability) an F_{ST} of 0.005 (Figure 3). In other words, by tripling the number of SNPs, we can halve the number of samples needed to detect a low expected F_{ST} (0.005). In contrast, as little as 40 to 70 samples per population were required for the 17 SNP markers to detect an F_{ST} of 0.01-0.02 with a similar level of statistical power (results not shown).

II. Population differentiation

SNP F_{ST} values ranged from 0 to 0.034 and microsatellite F_{ST} values ranged from 0.005 to 0.036 (Table 5). Consistent with the computer simulation results, both markers could detect an F_{ST} of > 0.03, finding significant structure between eastern Australia and California, with only the microsatellite markers also finding significant differentiation between western Australian and California, an F_{ST} < 0.03. Given the small sample sizes however, neither marker detected structure between the Australian humpback regions, observed using a larger microsatellite dataset in Schmitt *et al*. (In Press) (F_{ST} = 0.005, P = 0.001; D_{EST} = 0.031, P = 0.001).

Differentiation between regions was also significant using an alternative estimator of divergence, D_{EST}. D_{EST} values for the microsatellites were higher than for the SNPs which is consistent with the way D_{EST} partitions diversity. SNP D_{EST} values ranged from 0 to 0.015 and microsatellite D_{EST} values ranged from 0.009 to 0.124. For the eastern Australia/California region comparison, F_{ST} values for individual SNP loci ranged from 0 to 0.257 and D_{EST}, from 0 to 0.095 with five loci showing evidence of significant population structure (F_{ST} and D_{EST}, P < 0.05). For each of the microsatellite loci F_{ST} ranged from 0 to 0.133 and D_{EST} from 0 to 0.421 for individual loci with four of the ten loci showing evidence of significant population structure.

DISCUSSION

SNP markers are becoming increasingly easy to isolate from non-model organisms due to technological advances in high throughput sequencing, which means that identifying SNPs is a more viable and attractive alternative to microsatellite identification (Rosenberg *et al.* 2003). Some especially informative SNP markers (i.e. those that show the greatest allele frequency variation among populations) have even been known to exceed the average information content of microsatellite markers (Liu *et al.* 2005). The lower mutation rates with simple mutation models, the ability to easily combine data from different technologies and laboratories as well as the potential for long-term digital archiving, makes them an appealing marker for studies of whale population genetics. In this study we've shown that with sufficient sample sizes a small number of SNP markers, of varying quality and utility, can detect differentiation between both neighbouring and distant regions in the humpback whale (F_{ST} in the range of 0.005 to 0.04).

The POWSIM results indicated that for the small sample sizes used in this study (approximately 20 individuals per region), the 17 SNP markers had sufficient statistical power (at least an 80% probability) to detect differentiation between two populations at an F_{ST} of > 0.03 compared to an F_{ST} of > 0.01 for the ten microsatellites. Consistent with the power analysis, the AMOVA found significant structure between both Australian humpback whale regions and those from California using the ten microsatellite markers (EA and Cal, F_{ST} = 0.036; WA and Cal, F_{ST} = 0.029), but only between eastern Australia and California using the 17 SNPs (EA and Cal, F_{ST} = 0.033). A previous power simulation study found that 19 SNP loci ascertained from bowhead whales were able to detect an F_{ST} of 0.02 between two populations with a sample size of

55 (Morin *et al.* 2007b). In this study, only 40 samples were required to detect the same level of structure with our suite of 17 SNPs. Taking both studies into consideration, we suggest that a dataset of 20 informative SNPs and 50 samples per population will be adequate for detecting structure between distant humpback populations (i.e. between oceans and/or hemispheres), but not necessarily between more closely related populations.

Although the 17 SNPs have less statistical power to distinguish between closely related or demographically independent populations than the ten microsatellite markers due to the lower number of independent alleles, we found they could also detect weak population structure if larger sample sizes are used (Liu *et al.* 2005). Our results indicate that increasing the sample size results in significant gains in the statistical power of the SNP markers to detect population structure, as shown in other studies (Kalinowski 2005, Morin *et al.* 2009b). By increasing the sample size within each population to approximately 150 or more, the 17 SNP markers could detect structure between closely related populations (F_{ST} = 0.005) with a statistical power of greater than 0.8; equivalent to a dataset of 50 or more for the ten microsatellite loci. In fact, an increase in sample size from 50 to 100 provided a two-fold increase in the power of the 17 SNPs to detect weak differentiation, from 0.35 to 0.72 (proportion of significant χ^2 tests). This is in agreement with Kalinowski's (2005) and Morin's *et al*. (2009b) observation that increasing sample size is most beneficial for loci with low mutation rates, such as SNPs, and when F_{ST} is low (<0.01). Morin *et al.*(2009b) also found that increasing the sample size rather than increasing the number of SNP loci is likely to result in greater gains in statistical power. Our study supports this finding where the number of SNPs needed to be tripled (17 to 50), while the sample size only halved (150 to 75) to reliably detect an expected F_{ST} of 0.005. Given the ease of current methods of SNP discovery however, and the difficulty in collecting large sample sizes from Cetacean species, 50 SNPs combined with 75 samples would not be unreasonable to detect weak differentiation at this level of F_{ST}.

Although the power to resolve population differentiation is lower for individual SNPs than that of individual microsatellites because they are bi-allelic, we found that the set of SNPs examined here can provide sufficient power to detect population structure even when divergence between populations is very low (F_{ST} < 0.01). This study is one of the first to examine the utility of SNP markers in humpback whale population structure analysis, an area of research where microsatellites have, until now, been the most viable option for nuclear marker (Bourret *et al.* 2008). The challenges we face in understanding population structure in humpback whales remain considerable, including the collection of samples over many decades and the future need to combine data from researchers with access to samples in different oceanic regions. SNPs appear to offer the best option so far for overcoming those challenges and serve as a foundation for future regional and global collaborations. It is likely that the primers and sequence loci described here will also be useful for SNP discovery in a broader range of baleen species.

ACKNOWLEDGEMENTS

Research in Australia was funded by the Australian Marine Mammal Centre under permits from the Commonwealth of Australia and the states of Western Australia and New South Wales. Synthetic humpback DNA samples from California were provided by Scott Baker and Debbie Steel, and were prepared by Beth Slikas at Oregon State University.

LITERATURE CITED

Baker, C.S., L. Medranogonzalez, J. Calambokidis, A. Perry, F. Pichler, H. Rosenbaum, J.M. Straley, et al. 1998. Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. Molecular Ecology 7:695-707.

Barrett, J.C., B. Fry, J. Maller and M.J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265.

Bourret, V., M. Mace, M. Bonhomme and B. Crouau-Roy. 2008. Microsatellites in Cetaceans: An Overview. The Open Marine Biology Journal 2:38-42.

Brumfield, R.T., P. Beerli, D.A. Nickerson and S.V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology & Evolution 18:249-256.

Butler, J.M. 2005. Forensic DNA Typing: The Biology and Technology Behind STR Markers. Elsevier, New York.

Excoffier, L., G. Larval and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetic data analysis. Evolutionary Bioinformatics Online 1:47-50.

Freamo, H., P. O'Reilly, P.R. Berg, S. Lien and E.G. Boulding. 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Molecular Ecology Resources 11 (Suppl. 1):254-267.

Gagneux, P., C. Boesch and D.S. Woodruff. 1997. Microsatellite scoring errors associated with non-invasive genotyping based on nuclear DNA amplified from shed hair. Molecular Ecology 6:861-868.

Haasl, R.J. and B.A. Payseur. 2011. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity 106:158-171.

Hare, M.P. 2001. Prospects for a nuclear gene phylogeography. Trends in Ecology & Evolution 16:700-706.

Harpending, H.C., M.A. Batzer, M. Gurven, L.B. Jorde, A.R. Rogers and S.T. Sherry. 1998. Genetic traces of ancient demography. Proceedings of the National Academy of Sciences of the United States of America 95:1961-1967.

Hedrick, P. 2005. A standardized genetic differentiation measure. Evolution 59:1633-1638.

Hoffman, J.I., C.W. Matson, W. Amos, T.R. Loughlin and J.W. Bickham. 2006. Deep genetic subdivision within a continuously distributed and highly vagile marine mammal, the Steller's sea lion (*Eumetopias jubatus*). Molecular Ecology 15:2821-2832.

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

Jueterbock, A., P. Kraemer and G.D.J. Gerlach. 2010. *DEMEtics v0.8.0.* University of Oldenburg, Oldenburg, Germany.

Kalinowski, S.T. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? Heredity 94:33-36.

LaHood, E.S., P. Moran, J. Olsen, W.S. Grant and L.K. Park. 2002. Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. Molecular Ecology Notes 2:187-190.

Langmead, B., C. Trapnell, M. Pop and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25 doi: 10.11.86/gb2009-2010-2003-r2025.

Lao, O., K. van Duijn, P. Kersbergen, P. de Knijff and M. Kayser. 2006. Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. American Journal of Human Genetics 78:680-690.

Lessa, E. 1992. Rapid surveying of DNA sequence variation in natural populations. Molecular Biology & Evolution 9:323-330.

Li, S.Z., H.R. Wan, H.Y. Ji, K.Y. Zhou and G. Yang. 2009. SNP discovery based on CATS and genotyping in the finless porpoise (*Neophocaena phocaenoides*). Conservation Genetics 10:2013-2019.

Liu, N., L. Chen, S. Wang, C. Oh and H. Zhao. 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genetics 6:S26.

Macaubas, C., L. Jin, J. Hallmayer, A. Kimura and E. Mignot. 1997. The complex mutation pattern of a microsatellite. Genome Research 7:635-641.

Meirmans, P.G. and P.W. Hedrick. 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11:5-18.

Michalakis, Y. and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061-1064.

Morin, P.A., N.C. Aitken, N. Rubio-Cisneros, A.E. Dizon and S. Mesnick. 2007a. Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*). Molecular Ecology Notes 7:626-630.

Morin, P.A., B.L. Hancock and J.C. George. 2007b. Development and application of single nucleotide polymorphisms (SNPs) for bowhead whale population structure analysis. Paper SC/59/BRG8 presented to the IWC Scientific Committee, 2007 (unpublished). 7pp.

Morin, P.A., G. Luikart and R.K. Wayne. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution 19:208-216.

Morin, P.A., C. Manaster, S.L. Mesnick and R. Holland. 2009a. Normalization and binning of historical and multi-source microsatellite data: overcoming the problems of allele size shift with allelogram. Molecular Ecology Resources 9:1451-1455.

Morin, P.A., K.K. Martien and B.L. Taylor. 2009b. Assessing statistical power of SNPs for population structure and conservation studies. Molecular Ecology Resources 9:66-73.

Morin, P.A., V.L. Pease, B.L. Hancock, K.M. Robertson, C.W. Antolik and R.M. Huebinger. 2010. Characterization of 42 single nucleotide polymorphism (SNP) markers for the bowhead whale (*Balaena mysticetus*) for use in discriminating populations. Marine Mammal Science 26:716-732.

Palumbi, S.R. and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtdna of humpback whales. Molecular Biology & Evolution 11:426-435.

Parentage Testing Committee. 2003. Standards for Parentage Testing Laboratories. American Association of Blood Banks.

Pariset, L., I. Cappuccio, P.A. Marsan, S. Dunner, G. Luikart, P.R. England, G. Obexer-Ruff, et al. 2006. Assessment of population structure by single nucleotide polymorphisms (SNPs) in goat breeds. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 833:117-120.

Paschou, P., E. Ziv, E.G. Burchard, S. Choudhry, W. Rodriguez-Cintron, M.W. Mahoney and P. Drineas. 2007. PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genetics 3:1672-1686.

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Peakall, R., P.E. Smouse and D.R. Huff. 1995. Evolutionary implications of allozyme and RAPD Variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. Molecular Ecology 4:135-147.

Polanowski, A.M., N.T. Schmitt, M.C. Double, N. Gales and S.N. Jarman. 2011. TaqMan assays for genotyping 45 single nucleotide polymorphisms in the humpback whale nuclear genome. Conservation Genetics Resources DOI 10.1007/s12686-011-9424-5.

Rice, W.R. 1989. Analyzing tables of statistical tests. Evolution 43:223-225.

Rosenberg, N.A., L.M. Li, R. Ward and J.K. Pritchard. 2003. Informativeness of genetic markers for inference of ancestry. American Journal of Human Genetics 73:1402-1422.

Ryman, N. and S. Palm. 2006. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Molecular Ecology Notes 6:600-602.

Ryman, N., S. Palm, C. Andre, G.R. Carvalho, T.G. Dahlgren, P.E. Jorde, L. Laikre, et al. 2006. Power for detecting genetic divergence: differences between statistical methods and marker loci. Molecular Ecology 15:2031-2045.

Ryynanen, H.J., A. Tonteri, A. Vasemagi and C.R. Primmer. 2007. A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). Journal of Heredity 98:692-704.

Sacks, B.N. and S. Louie. 2008. Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. Molecular Ecology Resources 8:35-49.

Schmitt, N.T., M.C. Double, C.S. Baker, D. Steel, K.C.S. Jenner, M.-N.M. Jenner, D. Paton, et al. In Press. Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations. Marine Mammal Science.

Sevane, N., O. Cortes, D. Garcia, J. Canon and S. Dunner. 2010. New single nucleotide polymorphisms in Alectoris identified using chicken genome information allow Alectoris introgression detection. Molecular Ecology Resources 10:205-213.

Slade, R.W., C. Moritz, A. Heideman and P.T. Hale. 1993. Rapid assessment of single-copy nuclear DNA variation in divserse species. Molecular Ecology 2:359-373.

Smith, C.T., C.M. Elfstrom, L.W. Seeb and J.E. Seeb. 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology 14:4193-4203.

Smith, S., N. Aitken, C. Schwarz and P.A. Morin. 2004. Characterization of 15 single nucleotide polymorphism markers for chimpanzees (*Pan troglodytes*). Molecular Ecology Notes 4:348-351.

Vignal, A., D. Milan, M. SanCristobal and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34:275-305.

Winnepenninckx, B., T. Backeljau and R. Dewachter. 1993. Extraction of high molecular weight DNA from molluscs. Trends In Genetics 9:407-407.

TABLES AND FIGURES

Table 1. Results of a BLAST search in GenBank for sequenced intronic SNP loci

Locus	Accession No.	Species	Unique gene code	e-value	Max. identity %
SFRS15	BS000114.1	*Pan troglodytes*	intronic DNA	5.00E-50	75
TFRC	CU695181.11	*Sus scrofa*	intronic DNA	1.00E-51	80
MSL3	FP245417.7	*Sus scrofa*	intronic DNA	4.00E-36	84
LOC786990	NG_008397.1	*Homo sapiens*	FBXO11	1.00E-131	79
AKIRIN	NM_001101236.1	*Bos taurus*	AKIRIN1	3.00E-62	89
kin	BC149047.1	*Bos taurus*	intronic DNA	1.00E-76	88
UHMK	AL359699.12	*Homo sapiens*	intronic DNA	8.00E-29	80
ZRANB2	EZ461740.1	*Mustela putorius furo*	intronic DNA	3.00E-50	82
LOC513885	XM_001490761.1	*Equus caballus*	LOC100050763	5.00E-94	88
LOC781886	AC078860.19	*Homo sapiens*	intronic DNA	2.00E-43	75
ZC3H7B	AC192851.3	*Pan troglodytes*	intronic DNA	4.00E-17	74

Table 2. Primers designed and the number of SNPs identified for the 21 intron loci sequenced.

Locus	Aligned with cow Chromosome	Product size (bps)	No. of SNPs found	No. of individuals screened	F primer sequence	R primer sequence	T_A (^oC)
SFRS15	1	618	1	24	TGGCCCAACTGTTTCAGACAACTCAAGGTC	AACCTGAGCCATCACAGCATTGTCAATGGC	66
TFRC[c]	1	646	2	24	TATGTCGGAAAGGAGACTCTCTTGGAG	TCCGTGCTACGTCTAGACTAACAACCG	66
MSL3	1	840	1	8	TCTTAATCCATCCACCATCTCCTTACAAAGGTC	CAATGAGAGGCCTCGTCACCATCACGC	66
KIA	1	428	0	24	TTTACCATTTTCTTTTGAAACACTCAACTTGGG	AACCCCTGAGCTGTAGCCATCTCC	66
PHKA	1	976	0	8	GACTGTGATGAGAAGTTGTTTGACGATGCC	CAGATAGCACTTCGAGGATGTGCTTTGCAG	68
RPS6KA	1	725	0	8	TAACCATCCTTTTATTGTCAAGTTGCATTATGC	GAAAGTCCAATATAAGATACAACTTCCCTTCAG	64
UBXN7	1	454	0	24	GTGGACCCAAACAATGGAAGGGCC	GCAGGAAATGTATGTTCTAATAAGTGGTCTC	66
LOC786990[a]	1	640	1	8	CAGTTTCTCCTGAAGATACTGTTCAGC	TTTGCCAAAAAGAACAGCGTGTCCTAC	67
AKIRIN	3	746	1	8	ATGAGACAAAGACACTTCCCTACCTGC	TTTCATTATATAGTGTTAACAAGCTTAGTTGCAAAC	64
kin[b]	3	732	3	8	AATCAGAACAGACACGAGGTTCTCATG	AAGAACATGCTGATCATGTATTCTATATGTATGC	66
UHMK[c]	3	764	3	24	TTCTGCTTCCGGAGCCCGATACCCGTC	CATCATGAGGGTTATGTCCATGCAGACCTC	66
ZRANB2	3	796	1	8	GTTTAATACAGAAAAGTTTTGTGAATTGGAGAGGAG	TTAGAAGGTGGACTTTCAAGAAGTCTGAGGCAC	68
ARHGAP30	3	906	0	8	AGGAAGTCTTCCTGAGCGCCTATGATGACC	GGACCTGTTGGACCTGAACCAGGGTGG	66
MACF1	3	667	0	8	TATCATACACAGGCTGTTTGCCACCAAGAG	ATTGCTGGATGATGCCAGAAAGCGGGC	68
NCL	3	854	0	8	TCCAGCAAAGAAGATGATGGTTTCCCC	GCAAGATCATTTTTAGCAAAAAGGTCACTG	59
UAP1	3	660	0	8	ATTCTTCAGCGGGGAAAACTCATCTTCCCG	CCAGGGGCAGTTAATTAAGCCAGACAAACC	66
LOC513885	5	357	1	24	GAAACATTGGAAAGTCTATCGTAGGCGG	TGAATCAGCAGTGATAAAAGATGTTAAAGGTG	63
LOC781886	5	394	1	24	CTTGGACTTGATATTGGAGAATTGTCCAGTCTC	GCAAAAGCAGTGGCTAGTAAAGAAATAGG	64
ZC3H7B	5	384	1	8	CACCAGGGCATCTTCACTTTCCTCTGTGAG	TGCAGCGGGCATTGAAGCCAGCAGAGC	66
TSPO	5	839	0	8	GCAGCTGGCTCTGAACTGGGCATGGCC	TTGAGCATGGCCGCAAAGGCCAGCCAG	70
TMED7	5	345	0	24	TTCGGTACCTAAAATTAATTAGAGGACAGC	GCTTCAGCAATGAATTTTCTACTTTCACAC	63
Total[d]		13771	16	14.1			

[a]LOC786990 contained an indel of a T and a T/C SNP, so an assay could not be designed to isolate the SNP

[b]kin sequence could not be optimized, so was not used for genotyping

UHMK[c] and TFRC[c] - 1 SNP was not used for genotyping at each loci as it was too close to another SNP to provide enough flanking sequence for probe design

[d]Total represents the sum for sequence length, the number of SNPs found, but the average for the number of samples

Table 3. SNP genotyping and TaqMan primer sets

	Assay	F primer sequence	R primer sequence	Reporter 1 sequence (VIC dye)	Reporter 2 sequence (FAM dye)	SNP VIC	SNP FAM
Introns	SFRS15	GGGAAATGTGGGAAAGGTGAAAA	TTCCGCAGTTCTCTCCAAAGG	TGCTATTTCTCAGTGTGCTGT	CTATTTCTCAGCGTGCTGT	T	C
	LOC513885	AAGTGCAAGATGAAATTAAATTTAAAGCCGATA	CCGAGGTCGACATCTTTTCTTGATT	ATGTAACAGGAGTACTGC	TGTAACAGCAGTACTGC	G	C
	LOC781886	TGGGTGCCATTTAAAAACAAAACTTTAATCTTAT	GAGAAGGTTTTAATTATTCTTTTGAGGGTAAGAG	AAAAGCTTGAACTTTTT	AAAGCTTGAATTTTTTT	C	T
	UHMK1	GCCCACTTTCCATCCAACCA	TGTGGTAGCCAAGGACTCAAAATTT	CTCATTGCGATTTCAG	CTCATTGCAATTTCAG	G	A
	UHMK2	ACACTTTCCATACAACTGCTCATTTCT	AGCTCTTTGCAGAGTGTCCTTATTT	CTCTCTTGCCTGTCAGAC	CTCTCTTGCCTTTCAGAC	G	T
	TFRC	AATTTAAGAATGTTGAGAACCGGCTACT	CCCAATGCAGTTAGAGTTTACAAAGG	TTTACAATTTTAACATTAGAAGAT	CAATTTTAACATCAGAAGAT	T	C
	ZRANB2	GTTAATGAGGAACAGATTGAAGAAAGTTTTTTAAGT	AGTTAAACACCTCAAAGCAAATGTTCAC	AAGACCAACAGACATTGTGA	AGACCAACAGACGTTGTGA	A	G
	AKIRIN1	GCCCTGCCCTTGTTTTGG	GGTAGTCACTTGTTCCTGGAAGAG	CTGTAATGCGAGTGGCC	CTGTAATGCAAGTGGCC	G	A
	MSL3	GTGATGCAGAAGGTGGACAGA	GGACCGCCCCCTTCAG	CTCCCGGTGATTGG	TCCCGGTAATTGG	G	A
	ZC3H7B	CCCTCCACGGTCGATTCC	AGGTCGGGCCTCCAAAAC	CTCCACTGTGGCCTGT	CTCCACTGTAGCCTGT	G	A
Exons	Anon 1Y	GGGTGGTTCAGAGTGTCACTATCT	ACCAAGATGCAGCCAAAGCT	CTGACTTGCCATCTCC	CTGACTTGTCATCTCC	C	T
	HNLR	AGATCCTGACGCCGTTTACC	GGGCCGAAGTCCCAGAAAC	CATGGACACGGTCTCT	CATGGACACAGTCTCT	G	A
	Anon 2K	AGTCACCATCTCCCCACTACTC	AGAAGGAGAAACTTTGACTTTGGCA	CACTGTCTTTCTATTTGCT	CACTGTCTTTCTCTTTGCT	A	C
	TSPAN3	GTGGACCTGCTCAATTCACCTT	GGTGGAACCTATGCATAGTAGAAAACT	CCAAATCAGAGCAAGAC	CCAAATCAGAACAAGAC	G	A
	AHNAK	CCTCCTTCTATCTTTGGTCCTGAGA	GCTCTCAACTTGGATGCACCTAA	AAGTCCTCCGGCAAAC	AAGTCCTCCAGCAAAC	G	A
	P2YS	GGCTTCACTGACTGCGTTTT	TCCACTTGGTTCTTCAATAAGAAAATATTTTCATTTT	CTGGAATGACGGATTT	CTGGAATCACGGATTT	G	C
	BRCA1	CCACGCGGCCCTCTAG	CAGGCTGTCATCCTCTCCAAAA	CCCGTCACCTCCCACA	CCGTCACGTCCCACA	C	G
	ANON 3R	GGGCCCGAGAAAGCAGAT	GCATGGAGCAGGCAGATG	ATGGAGCTCACGTGCCGA	TGGAGCTCACATGCCGA	G	A
	ANON 4K	GAAGCGGACGTCCTTTCAC	CCTCTGCGTCCGAGCTTT	CCCCCGGGTGCGAG	TCCCCCTGGTGCGAG	G	T
	ANON 4S	CCGTGTGTGTTGTGGTTTTCAC	GGTGCTCAGCCACTCCTC	TCGCACCTGCTGCCCT	TCGCACCTCCTGCCCT	G	C

Table 4. Genetic diversity in humpback whales from Eastern and Western Australia and California genotyped at 17 SNP loci. N = number of genotyped individuals per locus and HW = deviation from Hardy-Weinberg equilibrium (p-value); significant at $P < 0.05$ (Standard errors in parentheses).

EA is eastern Australia; WA is western Australia; and Cal is California.

	N			Minor allele frequencies			Observed heterozygosity			Expected heterozygosity			HW (p-value)		
Locus	EA	WA	Cal	EA	WA	Cal	EA	WA	Cal	EA	WA	Cal	EA	WA	Cal
SFRS15	23	19	22	0.04	0.08	0.09	0.09	0.16	0.18	0.08	0.15	0.17	1.00	1.00	1.00
LOC781886	22	19	22	0.20	0.11	0.09	0.32	0.21	0.18	0.33	0.19	0.17	1.00	1.00	1.00
UHMK1	23	19	22	0.02	0.00	0.02	0.04	0.00	0.05	0.04	0.00	0.04	1.00	monomorphic	1.00
UHMK2	23	19	21	0.13	0.16	0.19	0.26	0.32	0.38	0.23	0.27	0.31	1.00	1.00	1.00
TFRC	23	19	21	0.04	0.08	0.14	0.09	0.16	0.29	0.08	0.15	0.24	1.00	1.00	1.00
ZRANB2	23	19	21	0.48	0.34	0.50	0.61	0.37	0.05	0.50	0.45	0.50	0.42	0.61	0.00
AKIRIN1	23	19	19	0.43	0.47	0.34	0.43	0.63	0.58	0.49	0.50	0.45	0.67	0.38	0.35
Anon 1Y	23	19	21	0.26	0.29	0.19	0.26	0.37	0.38	0.39	0.41	0.31	0.13	0.61	1.00
HNLR	23	19	22	0.13	0.13	0.36	0.17	0.16	0.36	0.23	0.23	0.46	0.31	0.26	0.37
Anon 2K	23	19	22	0.17	0.13	0.05	0.35	0.26	0.09	0.29	0.23	0.09	1.00	1.00	1.00
TSPAN3	23	19	22	0.28	0.21	0.00	0.30	0.42	0.00	0.41	0.33	0.00	0.30	0.54	monomorphic
AHNAK	23	19	21	0.46	0.32	0.45	0.74	0.32	0.52	0.50	0.43	0.50	0.04	0.29	1.00
P2YS	23	19	22	0.15	0.21	0.18	0.30	0.32	0.00	0.26	0.33	0.30	1.00	1.00	0.00
BRCA1	23	19	22	0.37	0.47	0.45	0.57	0.42	0.55	0.47	0.50	0.50	0.65	0.65	1.00
ANON 3R	23	19	22	0.41	0.47	0.45	0.57	0.63	0.73	0.48	0.50	0.50	0.67	0.38	0.08
ANON 4K	23	19	22	0.48	0.42	0.45	0.52	0.53	0.55	0.50	0.49	0.50	1.00	1.00	1.00
ANON 4S	23	19	20	0.11	0.18	0.30	0.22	0.37	0.50	0.19	0.30	0.42	1.00	1.00	0.62
All loci[a]	22.9 (*0.06*)	19.0 (*0.0*)	21.4 (*0.2*)	0.25 (*0.04*)	0.24 (*0.04*)	0.25 (*0.04*)	0.34 (*0.05*)	0.33 (*0.04*)	0.32 (*0.06*)	0.32 (*0.04*)	0.32 (*0.04*)	0.32 (*0.04*)	0.72 (*0.08*)	0.73 (*0.07*)	0.71 (*0.10*)

[a]All loci - represents the average across all loci.

Table 5. Population comparison among geographical areas using 17 SNP and ten microsatellite loci

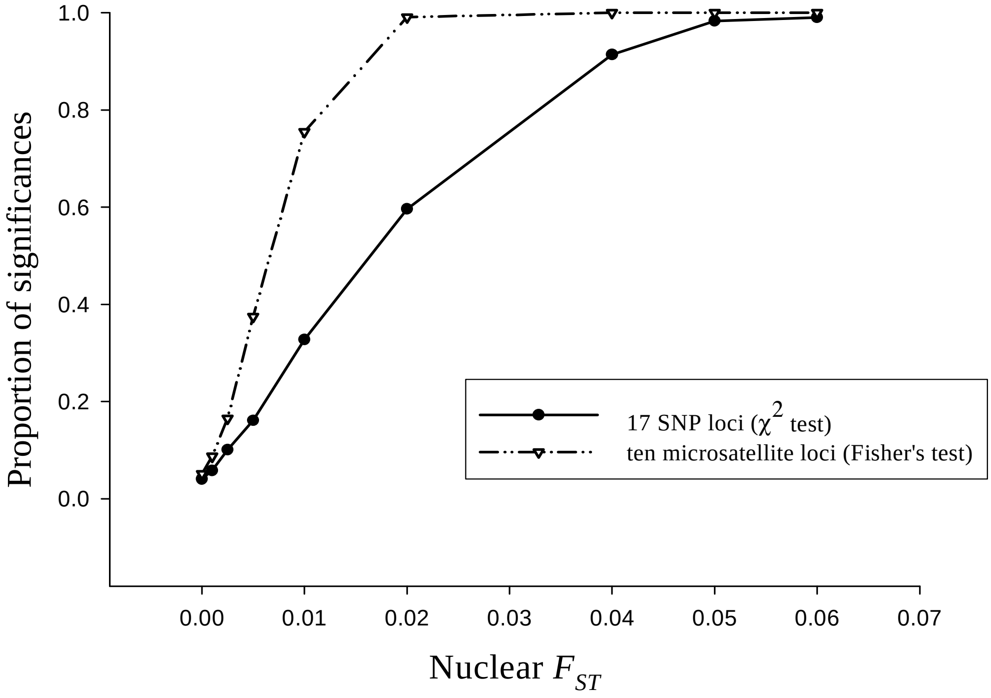
	EA		WA		Cal	
	SNPs	**microsats**	**SNPs**	**microsats**	**SNPs**	**microsats**
EA			0.000	0.028	0.015*	0.124*
WA	0.000	0.005			0.009	0.113*
Cal	0.034*	0.036*	0.018	0.029*		

F_{ST} values are presented in the lower left matrix and D_{EST} values are shown in the upper right. Statistically significant P-values based on 999 permutations of the data and after sequential Bonferroni correction are marked with an asterix (*P < 0.05).
EA is eastern Australia; WA is western Australia; and Cal is California.

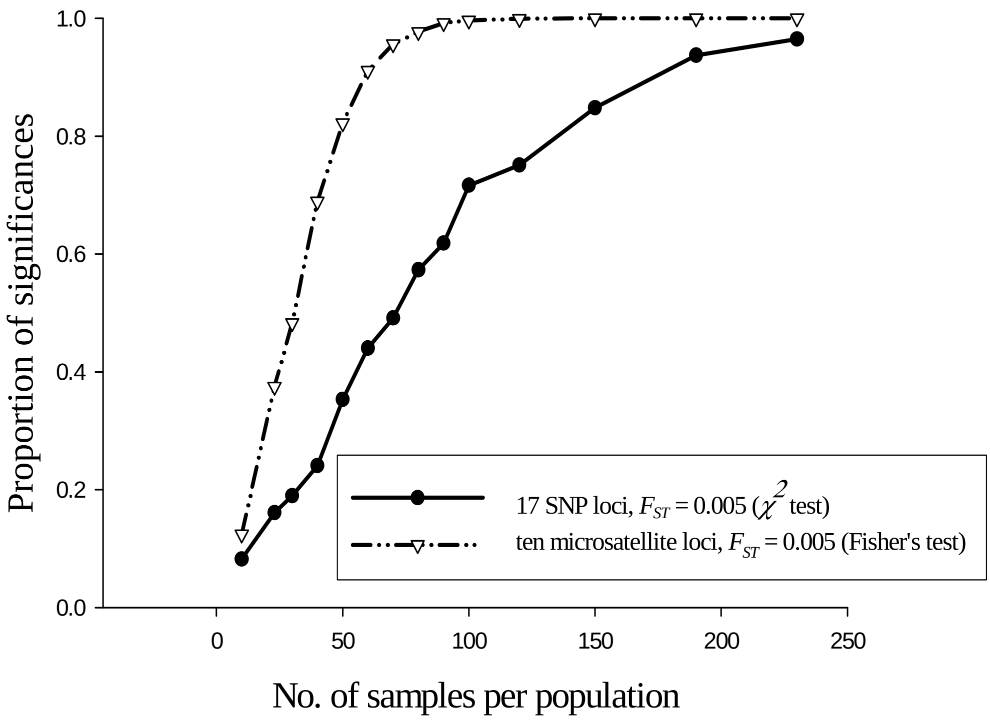
Figure 1. Simulated estimates of power (using the χ^2 test for SNPs and Fisher's test for microsatellites) to detect population differentiation between two populations when drawing a sample of n = 23 individuals from each at various true nuclear F_{ST}s for 17 SNP and ten microsatellite loci. At $F_{ST} = 0$ the proportion of significances represents the α-error.

Figure 2. Effect of sample size (number of individuals genotyped per population) on simulated estimates of power (using the χ^2 test for SNPs and Fisher's test for microsatellites) to detect population differentiation between two populations at the level of $F_{ST} = 0.005$ for the 17 SNP and ten microsatellite loci.

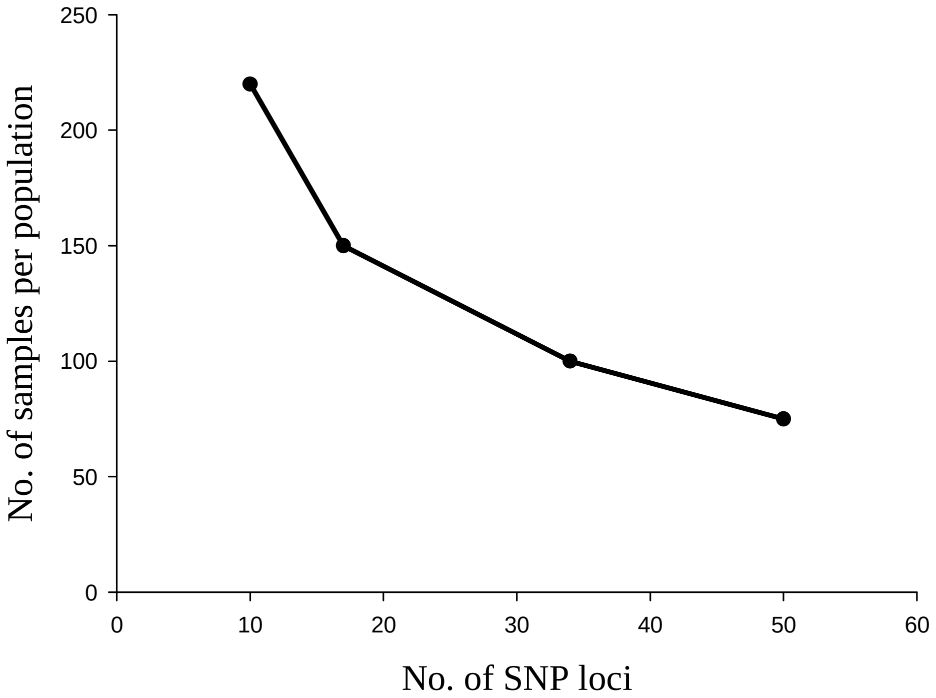
Figure 3. Effect of sample size per population on the statistical power (χ^2 test) to detect an expected level of $F_{ST} = 0.005$ for different numbers of SNP loci.



1.0
0.8
0.6
0.4
0.2
0.0
Proportion of significances
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Nuclear F_{ST}
17 SNP loci (χ^2 test)
ten microsatellite loci (Fisher's test)



1.0
0.8
0.6
0.4
0.2
0.0
Proportion of significances
0
50
100
150
200
250
No. of samples per population
17 SNP loci, F_{ST} = 0.005 (χ^2 test)
ten microsatellite loci, F_{ST} = 0.005 (Fisher's test)


No. of samples per population
250
200
150
100
50
0
0
10
20
30
40
50
60
No. of SNP loci

CONCLUSIONS

Utilising both mitochondrial and nuclear genetic markers, this thesis investigated the patterns of population structure among humpback whales that migrate along the east and west coast of Australia, and the neighbouring endangered populations of the South Pacific. Important in the development of demographic models that reconstruct the historical trajectory of population decline and recovery following the cessation of commercial whaling, this project investigated three structural parameters: population structure among putative breeding populations, the mixing of breeding populations on high latitude Antarctic feeding grounds and evidence for sex-specific migration along the eastern Australian migratory corridor. This knowledge will be crucial for the ongoing effort of the International Whaling Commission (IWC) in assessing the impact of whaling and the recovery of whale populations. We also report the discovery and utility of novel nuclear genetic markers (single nucleotide polymorphisms, SNPs). These markers hold promise for facilitating more effective multi-laboratory collaboration and are discussed under FUTURE RECOMMENDATIONS.

Among the Australian putative populations, both nuclear and mtDNA markers revealed low but significant differentiation. This pattern of low level differentiation is emerging as a characteristic of Southern Hemisphere humpback whale populations indicating extensive movement at least historically, if not presently. Contrary to the expectation of female philopatry and male-driven gene flow as displayed by many migratory marine vertebrates, there is limited evidence for strong sex biased dispersal among Australian humpback whales, with similar levels of genetic differentiation at both marker types evident between the sexes.

In understanding the impact of commercial whaling, an important gap in knowledge was estimating the mixing of low latitude breeding populations on Antarctic feeding grounds, particularly the endangered humpback whale populations of Oceania in the South Pacific. A mixed-stock analysis (MSA) of Antarctic samples collected south of eastern Australia and New Zealand using combined mtDNA and microsatellite datasets revealed substantial contributions from both eastern Australia (53.2%, 6.8% SE) and New Caledonia (43.7%, 5.5% SE) [with Oceania contributing 46.8% (5.9% SE)] but not western Australia. This result is significant in that non-genetic evidence based on photo-ID matching has suggested limited connectivity between the Antarctic feeding grounds where these samples were collected and New Caledonia (Oceania) but supports connections with eastern Australia (Constantine *et al.* 2011, Steel *et al.* 2011).

Before employing an MSA it was important to first evaluate the statistical power of the dataset and how we might go about improving the accuracy and precision of these analyses for future research. Despite the low differentiation among them, we found that all populations, whether Oceania was sub-divided (New Caledonia, Tonga, Cook Islands and French Polynesia) or grouped (minus the Cook Islands) using either combined mtDNA + microsatellite datasets or mtDNA alone performed close to or above the threshold for genetic identity (90%). This result therefore allows us to be relatively confident in the Antarctic feeding ground estimation. To increase our confidence, we found that removing the Cook Islands when the Oceania populations are treated separately and a moderate increase in breeding ground sample size had the greatest impact on the accuracy of an MSA.

While the coarse-scale spatial and temporal patterns of humpback whale seasonal movements between low latitude breeding grounds and high latitude feeding grounds are broadly understood, the assumption that the general pattern of migration for both sexes is similar is still under question. Along the eastern Australian migratory corridor, we found no compelling evidence for sex-specific migratory behaviour at three sampling locations, when the data was partitioned by sex and migration direction, i.e., males and females using different migratory routes, with no significant differences in the patterns of haplotype sharing, haplotype frequency or haplotype differentiation detected between the sexes. However, we did recover the same intriguing finding as a previous study (Valsecchi *et al.* 2010) of significant structure between males and females sampled off Stradbroke Island in 1992 at the nucleotide level, both overall and for whales travelling north towards the breeding grounds. We also found significant structure among all sampling locations at the haplotype level, irrespective of sex or migratory direction, and for northbound whales. These results suggest that humpback whale migration along eastern Australia may be more complex than previously thought, perhaps involving mixing with neighbouring populations, as well as social and age class temporal structuring. However, due to the small sample sizes used in some analyses, a final biological explanation for our results remains uncertain.

FUTURE RECOMMENDATIONS

Some recommendations have emerged from the biological and technical chapters of this thesis that warrant discussion here and further exploration for future humpback whale research.

Measuring gene flow when genetic differentiation is weak

The emerging evidence of low differentiation that appears to characterise populations in the Southern Hemisphere, presents significant challenges for reliable estimates of the magnitude of gene flow using many statistical approaches. Not only are estimates unreliable at low levels of genetic divergence, most population genetic inference methods are unable to discern between ongoing or recent historic migration because of the assumptions underlying the estimation of genetic divergence (see, Whitlock and McCauley 1999, Pearse and Crandall 2004, Pertoldi *et al.* 2007, Palsbøll *et al.* 2010). As a result of these challenges, clearly defining management units for humpback whales in the Southern Hemisphere becomes an important and difficult problem.

Novel approaches such as the kinship-based analyses of the spatio-temporal distribution of related individuals hold promise in being able to yield more reliable estimates of current migration rates at low levels of differentiation. However, formal estimation procedures applicable to a range of species life histories are required (Palsbøll *et al.* 2010). Other recent advances involve methods that allow the simultaneous analysis of genetic and non-genetic data to infer current migration and genetic drift using a Bayesian approach (see review in Gaggiotti and Foll 2010). However, presently these methods require further development as the spatial component is not taken into account explicitly.

Until we can obtain reliable estimates of current migration in systems where there is low differentiation between populations we suggest integrating information from multiple lines of evidence in Southern Hemisphere humpback whales. Capture-recapture models using photo-identification, genotype matching and the evolution and movement of song between populations, combined with estimates of population genetic structure may help delineate between current and historical gene flow and quantify the magnitude at a broad level.

Resolving population structure along the eastern Australian migratory corridor

Despite the lack of compelling genetic evidence for sex-specific migratory behaviour, our data provided potential evidence for temporal structure along the eastern Australian migratory corridor. The significant structure detected between three sampling locations at the haplotype level, and between at least two sampling locations for three other statistics, suggests there could at least be temporal differences and potentially more than one population using the corridor at any one time. Possible biological explanations for our study include the sporadic sharing of the eastern Australian migratory corridor by males and females from neighbouring breeding grounds during migration, social structuring of the migratory population or even commercial whaling leaving a signal on the genetic composition of age classes. Resolving these structural complexities is important as they may significantly impact our estimates of abundance for the eastern Australian population and the modelling of their recovery from historical overexploitation, particularly if there is more than one population using the migratory corridor. There may also be broader consequences if these structural complexities occur along other migratory routes in the Southern Hemisphere.

At this stage we have insufficient samples from eastern Australian humpback whales for a quantitative evaluation of temporal genetic structure along the migration from feeding grounds to breeding grounds, as well as the effects of age and social structure on temporal genetic shifts. We also lack data on how the effective size of the population and the patterns of intermixing may have changed as a result of commercial whaling, particularly with the eastern Australian population increasing at a more rapid rate (10 to 11% per annum) than neighbouring populations which are still considered endangered (Childerhouse *et al.* 2008, Noad *et al.* 2011). Resolving the potential complexities along the eastern Australian migratory corridor will involve improving the geographical and temporal coverage of humpback whale samples from the Southern Ocean through to the breeding grounds around the Great Barrier Reef, where we are yet to obtain any samples. We also need to maintain focus on possible mixing with neighbouring populations, age and social structure, and better understanding the influence of historical overexploitation and recovery on temporal structure.

Enhancing inferences of population structure and improving the resolution of mixed-stock analyses – mitogenomics?

Defining management units for humpback whales within eastern Australia and Oceania has been a contentious issue within the International Whaling Commission due, in part, to the limitations associated with the weak genetic differentiation between them. Accurate assessments of population structure and migratory connectivity are important for effective management and for modelling the recovery of these putative populations. Moreover, the weak genetic divergence among populations introduces an element of uncertainty into estimates of population contributions on the Antarctic feeding grounds (see Chapter 2), even if divergence is significant based on haplotype and microsatellite allele frequency differences.

To date, short fragments of the mitochondrial control region have been the marker of choice in analyses of population structure among humpback whales due to its highly discriminatory and uniparental inheritance which are expected to result in a larger genetic drift compared to nuclear loci (e.g. Avise 1995, Sunnucks 2000). Nuclear markers have generally detected equivalent or considerably less structure than that inferred using mitochondrial markers (see Chapters 1 and 2). As mtDNA control region haplotypes are widely shared across Australia and the South Pacific, expanded screening of the mitochondrial genome may benefit analyses of genetic structure among populations. MtDNA analysis beyond established control region fragments has improved the resolution of intraspecific phylogeography and population structure of several marine taxa. Phylogeographic analysis of whole mitochondrial genome sequence variation in killer whales (*Orcinus orca*) provided strong support for species status of the ecotypes (Morin *et al.* 2010), whereas an earlier analysis based on shorter sequences failed to resolve these relationships because of the limited polymorphism (Hoelzel *et al.* 2002). Another study on green turtles (*Chelonia midas*) demonstrated the utility of mitogenomic SNPs for detecting cryptic structure among populations that are marked by extensive sharing of common haplotypes based on <1 kb of the mitogenome, a similar scenario to Southern Hemisphere humpback whales (Shamblin *et al.* 2012).

Mitogenomic sequences may therefore increase the statistical power of our humpback whale genetic data to detect cryptic structure among putative populations as well as increasing certainty in mixed-stock analyses on the feeding grounds and other areas of potential population mixing.

SNPs – the marker of the future for long-term collaborative cetacean research?

Although microsatellites have been the nuclear marker of choice in studies of cetacean population genetics, the research would greatly benefit from the use of single nucleotide polymorphisms (SNPs) as scoring is universally comparable and mutation rates and patterns of are well described by simple mutation models making them suitable for long-term, collaborative studies of globally distributed species.

In the fourth, technical chapter on the development and application of SNP markers for humpback whale population genetics, we developed 17 intronic and exonic SNPs using a targeted gene approach that had sufficient statistical power to detect weak structure that is typical among some populations (F_{ST} = 0.005) with sample sizes of >150. Given the ease of our methods of SNP discovery and the difficulty in collecting large sample sizes from Cetacean species, we recommended that 50 of these informative SNPs combined with 75 samples would not be unreasonable to detect weak differentiation at this level of F_{ST}. Since writing the review on SNP discovery and application however, high throughput sequencing technology has advanced even further and is now considered a highly effective platform for SNP discovery. A particularly efficient and cost-effective protocol based on the sequencing of large sets of restriction fragments that holds great promise for future SNP discovery in humpback whales and cetaceans in general, is termed "restriction-site associated DNA" (RAD) (Miller *et al.* 2007), in combination with the Illumina Genome Analyzer sequencing device. With the capability of generating thousands of informative SNP markers and the sequenced genome of the common dolphin (*Tursiops truncatus*) now available as a reference for the alignment of contigs, there is much potential in discovering a combination of markers for humpback whales that provide better quality information than is possible from currently employed marker systems.

As well as being useful in helping to resolve cryptic structure among some humpback whale populations, the universal comparability of SNPs may allow us to create a centralised genetic database for cetaceans, combining data from researchers with access to samples collected over many decades in different oceanic regions. SNP databases for human genetics studies and forensics are already well established with free access to a substantial proportion of the 18 million SNP loci currently characterised in the human genome.

Although there are still problems associated with SNP markers that present challenges for their broad use in humpback whale population genetic analyses, particularly when dealing with large numbers of loci (e.g. ascertainment bias, linkage between loci, and neutrality), the most effective application may be in combining them with the suite of genetic markers we already have available (e.g. sperm whales Mesnick *et al.* 2011).

The findings of this thesis provide new information on the population structure and distribution of the humpback whales of Australia and the South Pacific. Given the weak genetic differentiation that appears to characterise whales from this region, determining accurate structure hypotheses for modelling the recovery of the species will continue to be contentious. With the new era of population genomics and high throughput sequencing technologies in the development of new, informative markers we can substantially increase the statistical power of our data. Exploring new and innovative ways of combining our genetic and non-genetic data in delineating patterns and levels of gene flow will also contribute substantially to our understanding of the dynamics of humpback whales from these two regions.

LITERATURE CITED

Avise, J.C. 1995. Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. Conservation Biology 9:686-690.

Childerhouse, S., J. Jackson, C. Baker, N. Gales, P. Clapham and R.L. Brownell Jr. 2008. *Megaptera novaeangliae* (Oceania subpopulation) IUCN 2012. IUCN Red List of Threatened Species. Version 2012.1.

Constantine, R., J. Allen, P. Beeman, D. Burns, J. Charrassin, S. Childerhouse, M.C. Double, et al. 2011. Comprehensive photo-identification matching of Antarctic Area V humpback whales. Paper SC/63/SH16 presented at the IWC Scientific Committee, Tromso, Norway.

Gaggiotti, O.E. and M. Foll. 2010. Quantifying population structure using the F-model. Molecular Ecology Resources 10:821-830.

Hoelzel, A.R., A. Natoli, M.E. Dahlheim, C. Olavarria, R.W. Baird and N.A. Black. 2002. Low worldwide genetic diversity in the killer whale (*Orcinus orca*): implications for demographic history. Proceedings of the Royal Society of London Series B-Biological Sciences 269:1467-1473.

Mesnick, S.L., B.L. Taylor, F.I. Archer, K.K. Martien, S.E. Trevino, B.L. Hancock-Hanser, S.C.M. Medina, et al. 2011. Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. Molecular Ecology Resources 11:278-298.

Miller, M., J. Dunham, A. Amores, W. Cresko and E. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research 17:240-248.

Morin, P.A., F.I. Archer, A.D. Foote, J. Vilstrup, E.E. Allen, P. Wade, J. Durban, et al. 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Research 20:908-916.

Noad, M.J., R.A. Dunlop, D. Paton and D.H. Cato. 2011. Absolute and relative abundance estimates of Australian east coast humpback whales (*Megaptera novaeangliae*). Journal of Cetacean Research and Management (special issue 3):243-252.

Palsbøll, P.J., M.Z. Peery and M. Berube. 2010. Detecting populations in the 'ambiguous' zone: kinship-based estimation of population structure at low genetic divergence. Molecular Ecology Resources 10:797-805.

Pearse, D.E. and K.A. Crandall. 2004. Beyond F-ST: analysis of population genetic data for conservation. Conservation Genetics 5:585-602.

Pertoldi, C., R. Bijlsma and V. Loeschcke. 2007. Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. Biodiversity and Conservation 16:4147-4163.

Shamblin, B.M., K.A. Bjorndal, A.B. Bolten, Z.M. Hillis-Starr, I.A.N. Lundgren, E. Naro-Maciel and C.J. Nairn. 2012. Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. Molecular Ecology 21:2330-2340.

Steel, D., N.T. Schmitt, M. Anderson, D. Burns, S. Childerhouse, R. Constantine, T. Franklin, et al. 2011. Initial genotype matching of humpback whales from the 2010 Australia/New Zealand Antarctic Whale Expedition (Area V) to Australia and the South Pacific. Paper SC/63/SH10 presented at the IWC Scientific Committee, Trömso, Norway.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in Ecology & Evolution 15:199-203.

Valsecchi, E., P.J. Corkeron, P. Galli, W. Sherwin and G. Bertorelle. 2010. Genetic evidence for sex-specific migratory behaviour in western South-Pacific humpback whales. Marine Ecology Progress Series 398:275-286.

Whitlock, M.C. and D.E. McCauley. 1999. Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). Heredity 82:117-125.

Appendix: Development and application of SNP's for humpback whale population genetics (literature review)

In less than half a century, molecular markers revealing polymorphisms at the DNA level have totally changed our view of nature, and in the process they have evolved themselves. Nonetheless, the current repertoire of genetic methodologies remain inadequate for answering many questions and there are significant technological and analytical limitations.

Over the last decade, nuclear microsatellite loci and mitochondrial DNA (mtDNA) sequences have been the tools of choice for population genetic studies. Both types of molecular marker represent rapidly evolving DNA sequences that are informative for answering population-level questions. However, while useful and indeed optimal for some applications, the high information content, a result of high mutation rates, comes at a price. Analytical issues include the limitations of depending on a single locus (mtDNA, Gagneux *et al.* 1997, Harpending *et al.* 1998, Hare 2001), high and variable mutation rates that are difficult to predict for data analysis (microsatellites, Excoffier and Yang 1999, Balloux 2002) and limitations due to sample size (B-Rao 2001). Technical problems include nuclear inserts of mitochondrial DNA (numts, Bensasson *et al.* 2001, Dunshea *et al.* 2008), and microsatellite null alleles, and allelic dropout (Navidi *et al.* 1992, Callen *et al.* 1993, Taberlet *et al.* 1996, Gagneux *et al.* 1997, Morin *et al.* 2001). Microsatellite markers can be particularly limited in long-term population studies that require addition of data over time and across multiple labs, due to inconsistencies in allele size calling (LaHood *et al.* 2002, Vignal *et al.* 2002, Hoffman *et al.* 2006).

Ideally, molecular markers for population studies of non-model organisms would possess three properties. First, the markers should be abundant and distributed widely across the genome to avoid biases associated with single locus analysis. Second, well-understood and well-characterized models of evolution should apply to facilitate analysis and interpretation. Finally, technical applications must allow data acquisition from many loci scored in large population samples, and the data must be comparable across laboratories using different genotype scoring methods or technologies (Sunnucks 2000). Genetic analyses of wide-ranging, elusive species like the humpback whale (*Megaptera novaeangliae*) will benefit from using markers that allow datasets to build over a period of decades and that allow sample size augmentation, both spatially and temporally.

Single nucleotide polymorphisms (SNPs) are rapidly becoming the tool of choice for assessing genetic variation in wild populations as they have many of the characteristics of an ideal marker (Vignal *et al.* 2002, Brumfield *et al.* 2003, Morin *et al.* 2004): 1) They are frequent, abundant and easy to ascertain in many non-model organisms (Aitken *et al.* 2004, Seddon *et al.* 2005, Cappuccio *et al.* 2006, Morin and McCarthy 2007). 2) The rate and patterns of SNP mutation have been characterized extensively in several genomes and are well described by simple mutation models (Nachman and Crowell 2000, Ebersberger *et al.* 2002, Silva and Kondrashov 2002). 3) SNP genotypes are based on detection of DNA sequence nucleotide differences rather than PCR product size differences (microsatellites), so that genotype data are universally comparable and portable, eliminating the need for common controls (Morin *et al.* 2007b). Hence, SNP studies can be replicated, performed in parallel across several laboratories, and added to as samples become available without the need to calibrate results at each step in the process. 4) SNP genotyping technologies vary widely, ranging from simple to highly multiplexed for high throughput genotyping, allowing a broad choice of systems (Vignal *et al.* 2002, Morin *et al.* 2007b).

In spite of these obvious advantages, and the increasing use of SNP's in human and model organism studies, they have not been employed frequently in studies of non-model organisms. This is primarily due to technical and analytical issues in SNP isolation as well as cost-effectiveness and efficiency of genotype production in relatively unknown genomes (Vignal *et al.* 2002, Aitken *et al.* 2004). In this review I examine the feasibility of developing and applying SNP's to the genetic analyses of humpback whales, focusing specifically on new innovating methods of SNP discovery and genotyping that endeavour to reduce cost and effort, thus enabling the extraction of better quality information than is possible from currently employed marker systems.

A SNP is a single nucleotide variation at a specific location on the genome that is by definition found in more than 1% of the population. Although in principle, at each position of a sequence stretch, any of the four possible nucleotide bases can be present, SNPs are usually biallelic in nature and co-dominant (Vignal *et al.* 2002, Kim and Misra 2007). Being the most abundant genetic variant within a genome (coding and non-coding regions, typically every 300-1000 bps in most genomes), the increasing popularity of SNPs has arisen from the recent need for very high densities of genetic markers for the studies of multifactorial diseases, and the recent progress in polymorphism detection and genotyping techniques (Vignal *et al.* 2002). The use of SNP markers involves two principle steps: locus discovery (ascertainment) and genotyping.

SNP DISCOVERY

SNP discovery is the process of finding the polymorphic sites in the genome of the species and populations of interest. The most well known strategies for SNP discovery utilized in the human genome project are performed *in silico*, meaning genomic information from multiple individuals in public databases are screened for the identification of putative polymorphisms. These strategies include: expressed sequence tags for polymorphic sites and reduced representation shotgun sequencing (Schlotterer 2004).

Expressed Sequence Tags (EST)

The simplest approach using public databases, where a *de novo* whole-genome sequencing project is not yet feasible in a given organism for reasons of cost or technical difficulty, is a screen of expressed sequence tag databases which focuses on protein-coding sequence by sequencing only mRNA (Adams 1991, Hudson 2008). While an EST project does not produce more than a fraction of the information available from a whole-genome sequence, it can produce a 'tag' of sequence from the protein coding region of most genes. Because the majority of these EST libraries have been obtained from different individuals, assembly of overlapping sequences (or 'tags') for the same region can lead to the identification of new SNPs. A significant risk in such an analysis is that many sequence variations are the result of poor quality sequence data typically found in single-pass EST datasets. Therefore, the success of any strategy using EST data sets hinges on strategies used to cull sequence errors from the candidate pool (Picoult-Newberg *et al.* 1999).

Reduced representation shotgun sequencing (RRS)

RRS is a new and more comprehensive approach, used to isolate a very high number of SNPs in humans (Altshuler *et al.* 2000). Here DNA from different individuals are mixed together and plasmid libraries composed of a reduced representation of these genomes are produced by using a subset of restriction fragments purified by agarose gel electrophoresis. A 2-5 fold shotgun sequencing of the libraries is performed and aligned overlapping sequences are screened for polymorphism.

Both strategies are at risk of false positives due to the alignment of sequences from repeated loci (Vignal *et al.* 2002). This can be partially overcome where species databases of repeated elements are available, that can be used to filter the sequence reads prior to alignment. However, the case of duplicated loci will always remain difficult to manage.

For most non-model organisms, in the absence of databases of comparative sequences, SNP's have to be found through laboratory screening (e.g. sequencing) of segments of the genome from multiple individuals (Morin *et al.* 2004). The most common screening strategies to maximize the number of SNP loci identified across a genome are: 1) designing primers from conserved sequences of related species that are available online or 2) sequencing anonymous nuclear loci, either through cloning or PCR methods, such as amplified fragment length polymorphisms (AFLPs) (Brumfield *et al.* 2003). These approaches are applicable to any organism and produce large numbers of DNA fragments that can be readily sequenced for SNP screening with relatively little investment.

AFLPs – a random sequence approach

The amplified fragment length polymorphism method of SNP discovery is a PCR-based technique that has been successfully applied to a wide range of organisms with a broad application to population and evolutionary genetics (McLenachan *et al.* 2000, Bensch *et al.* 2002, Nicod and Largiader 2003). The technique is based on the selective PCR of restriction fragments from a total digest of genomic DNA involving three steps: 1) restriction of the DNA and ligation of oligonucleotide adapters; 2) selective amplification of sets of restriction fragments and 3) gel analysis of amplified fragments. SNP markers are then isolated from the direct sequencing of distinct electrophoresis bands which highlight several homologous copies of a particular DNA fragment. The efficiency of this strategy in SNP isolation will depend critically on the proportion of AFLP bands that represent unique DNA fragments, SNP density in the organism under study and the quality of the DNA utilized (the effectiveness of AFLP is reduced when DNA quality is poor) (McLenachan *et al.* 2000, Nicod and Largiader 2003). Nonetheless, in terms of cost, isolation of a couple of hundred candidate SNPs involves about the same expense and time effort as is usually invested to isolate microsatellite markers for population genetics studies (Nicod and Largiader 2003).

The targeted gene approach

As a way of complimenting existing mtDNA and microsatellite markers for a given species, allowing the partitioning of maternal and biparental components of dispersal and population structure, a number of mammalian studies have used a target gene approach to detect SNPs and nuclear intron sequences (Palumbi and Baker 1994, Aitken *et al.* 2004, Seddon *et al.* 2005, Cappuccio *et al.* 2006, Morin *et al.* 2007a). The targeted gene or genomic region approach exploits the fact that there are often conserved regions of genes from multiple species (e.g. mouse and human) from which PCR primers can be designed to amplify a less conserved region in related species (e.g. an intron, or microsatellite). These primers have been termed 'comparative anchor tagged sequences' ('CATS'; (Lyons 1997) or 'exon priming intron crossing' ('EPIC'; (Palumbi and Baker 1994). The advantages of this method include the wide availability of primers, knowledge of the gene ortholog which contain the SNPs, which may be useful in detecting selection on ecologically important genes (Palumbi and Baker 1994, Primmer *et al.* 2002, Aitken *et al.* 2004) and potentially broad taxonomic utility with less per-sequence initial effort to find SNP loci than might be required using a random sequence approach (Morin *et al.* 2004).

Introns are a common source of nuclear sequence data as they evolve with fewer evolutionary constraints relative to coding sequence (Palumbi and Baker 1994, Prychitko 1997, Friesen 1999, Hare 2003). The use of intron sequences as a source of variation where the more highly conserved regions (i.e. exons) of the genes are used for primer design, has been exploited extensively to date in both phylogenetic (e.g. (Oakley 1999, DeBry 2001) and population genetic studies (e.g.(Lessa 1992, Bierne 2000). As with screening methods for model species, there are similarly technical and conceptual limitations to this approach that may affect SNP identification. Foremost is the increased risk of amplifying both paralogs of the same locus which can result in an incorrect inference of genotypes, and a biased inference of historical events resulting from the effects of selection acting on the associated genes (Aitken *et al.* 2004, Ryynanen and Primmer 2006). A new method termed intron-primed exon-crossing (IPEC) was developed by Ryynanen and Primmer (2006) to circumvent this problem in species bearing putative duplicated genomic fragments or an entire duplicated genome, whereby primers were designed in more variable intronic regions of salmonid genes to bind to only one of the duplicated loci. The proportion of screened loci that revealed polymorphism was around six times higher with the IPEC than the EPIC method, suggesting that less effort is needed to yield the same number of SNPs than with the EPIC (or CATS) method. The success of this technique of SNP isolation in humpback whales would depend upon the availability of more than one sequence copy of the particular gene from related species in the databases and knowledge on the extent of duplicated genes in different species (would also need to ascertain costs). (Baker *et al.* 2006) found evidence of gene duplication as well as high allelic and nucleotide diversity for at least two DQB loci in the humpback whale and preliminary evidence for three in the right whale, supporting the feasibility of the IPEC strategy for detecting SNPs in baleen species.

Generating novel SNP loci using the targeted gene approach can require significant effort, as a large investment in CATS primers is necessary (≥200 primer pairs, ~US$25 a pair)

(Morin *et al.* 2004). Although the initial investment in primers should ideally be spread out over many species, making the relative cost of developing markers for several species substantially lower, this project focuses specifically on developing SNPs for humpback whale population analysis. A comparison of SNP frequencies across a diverse range of taxa suggests that researchers can expect, on average, to find SNPs in over 50% of loci. Hence, 100 loci (each of 500-800 bp in length) can be sequenced for ≤US$6000 with SNP assays assembled for ~US$3750 (Brumfield *et al.* 2003, Morin *et al.* 2004). By comparison, discovery of 15 new microsatellites, using a commercially generated enriched library, would cost ~US$12000.

Whichever sequence generation method is employed, SNP detection requires multiple sample sequences to be generated from each locus (Aitken *et al.* 2004). The last "in silico" step of identifying the SNPs in the sequence traces, has greatly benefited from the development of specialized software estimating the quality of base calling such as PHRED (Ewing 1998a, Ewing 1998b) and of other programs using this quality assessment for polymorphism detection such as POLYPHRED (Nickerson 1997) or POLYBAYES (Marth 1999). These programs assist in SNP detection, either directly from all amplified samples, or on a subset of samples subsequent to high throughput (but potentially less accurate) methods of mutation screening (Wolford 2000, Brumfield *et al.* 2003, Aitken *et al.* 2004).

SNP Discovery in whales

Only four studies to date have employed the use of SNPs for the genetic analysis of whales: one for sperm whale population studies (Morin *et al.* 2007a) and three using historical samples to supplement existing tissue collections for the genetic analyses of Bering-Chukchi-Beaufort (BCB) bowhead whales (Givens 2007, Morin *et al.* 2007b, Morin and McCarthy 2007).

Morin *et al.* (2007a) utilized two hundred and two comparative anchor tagged sequence (CATS) primer pairs to identify 18 new SNP markers from a pool containing 20 individual sperm whales representing a broad geographical distribution and six DNA samples from one individual. Although their intent was to use the pooled sample to identify SNPs common in some populations but not others, most SNPs were ascertained using the six individual samples, and the pooled samples served only to confirm the presence of some SNPs. High quality sequence data was screened for SNPs using POLYPHRED.

Givens (2007), Morin *et al.* (2007b) and Morin and McCarthy (2007) identified SNPs by screening random genomic DNA sequences using primers designed from existing cloned sequences from a single bowhead whale. Sequencing was performed using highly multiplexed PCR of small (<120bp) products, a method previously developed for sequencing of mtDNA and single copy nuclear genes from ancient DNA samples (Krause 2006, Rompler 2006a, 2006b). The estimated cost of this technique based on an NPRB proposal summary in 2005, to produce about 30-35 independent SNPs for population analysis, was US$8079.

SNP GENOTYPING

Genotyping typically involves the generation of allele-specific products for SNPs of interest followed by their separation and detection for genotype determination (Vignal *et al.* 2002, Kim and Misra 2007). Unlike microsatellite markers that have a standard procedure for genotyping involving PCR and size determination of the amplified fragment, there exist at least 20 different SNP genotyping methods today to meet the cost, throughput and equipment needs of each laboratory (Tsuchihashi 2002). Technologies can range from simple and standard (such as electrophoresis-based systems) to highly multiplexed for high throughput genotyping (e.g. microarrays Morin *et al.* 2007b). Most current genotyping technologies with only a few exceptions rely on pre-amplification of the SNP-containing genomic region by PCR, to introduce specificity and increase the number of molecules for detection following allelic discrimination (Kim and Misra 2007).

Allele discrimination strategies

Allele discrimination is performed using allele-specific biochemical reactions of which there are four popular methods: primer extension (nucleotide incorporation), hybridisation, ligation and enzymatic cleavage.

I. Primer Extension (Advantage: Accurate genotyping, Disadvantage: Similar reagents as in PCR)

These approaches involve allele-specific incorporation of nucleotides in primer extension reaction with a DNA template, utilizing enzyme specificity to achieve allelic discrimination. The primer extension reaction is described as robust, allowing specific genotyping of most SNPs at similar reaction conditions (Syvanen 2001) This feature is advantageous for high throughput applications as the effort required for assay design and optimization is minimized. Consequently, primer extension is gaining acceptance as the reaction principle of choice for high throughput genotyping of SNPs, and has been adapted to various assay formats, detection strategies and technology platforms (Syvanen 2001, Morin *et al.* 2004). Primer extension assays either use a common primer for detecting both alleles (single base extension or SBE) or specific primers for detecting each allele (allele specific primer extension or ASPE) (Kim and Misra 2007).

With single base extensions, a primer is designed to anneal with its 3' end adjacent to a SNP site and extended with nucleotides by polymerase enzyme (Sokolov 1990). The identity of the extended base is determined either by fluorescent labelling or mass to reveal the SNP genotype. Primer selection and assay design are simple and allow for the simultaneous selection of multiple SNPs, thereby reducing costs and improving throughput (Vignal *et al.* 2002). Hence a large number of commercial systems employ SBE for SNP genotyping (e.g. mass spectrometry: The PinPoint assay, MassEXTEND, SPC-SBE, and GOOD assay; fluorescent labelling: SNaPshot, SNPstream and Microarray primer extension, see (Syvanen 2001, Tsuchihashi 2002, Morin *et al.* 2004, Kim and Misra 2007) for comparisons).

Allele-specific primer extension (ASPE) involves PCR amplification of genomic DNA using allele-specific primers that are labelled differently and a common reverse primer. The extended PCR product is generated only when the allele-specific primer binds perfectly to a SNP site and can be detected by fluorescence to determine SNP genotype (Ugozzoli 1992, Kim and Misra 2007). However, in practice, the PCR conditions can be difficult to set up and reliability is low (Vignal *et al.* 2002). Examples include the Tag-It approach (Bortolin 2004) and performing the primer extension directly on microarrays (Pastinen 2000).

II. *Hybridisation (Advantage: Widely used, Disadvantage: Limited genotype discrimination)*

Hybridisation approaches use differences in thermal stability of double stranded DNA to distinguish between perfectly matched and mismatched target-probe pairs for achieving allelic discrimination (Kim and Misra 2007). Although conceptually simple, hybridisation techniques are error prone and need carefully designed probes and protocols so that hybridisation occurs only between a perfectly complimentary probe and target (Vignal *et al.* 2002). The effectiveness in allele differentiation will generally depend upon the length and sequence of the probe, location of the SNP in the probe, and hybridisation conditions. These parameters must be determined empirically and separately for each SNP. Consequently, there is no single set of reaction conditions that would be optimal for genotyping all SNPs, which makes the design of multiplex assays based on hybridisation an almost impossible task (Syvanen 2001).

One approach to circumvent the problem of assay design is to carry out multiplex allele specific oligonucleotide (ASO) hybridisation reactions on microarrays that carry very high probe densities for each SNP to be analysed. One example, GeneChip array technology

(Affymetrix, CA), uses a computer algorithm to interpret the complex fluorescence patterns formed by the multiple probes and to assign the genotypes of each SNP (Kennedy 2003). However, distinguishing between heterozygous and homozygous SNP genotypes can be a problem with this technique.

By far the most widely used ASO hybridisation methods distinguish between the SNP alleles in real time during PCR in homogenous, solution-phase hybridisation reactions with fluorescence detection (Syvanen 2001). The TaqMan (Applied Biosystems) or Molecular Beacon probes (high expense for probes), which were originally designed for quantitative PCR analysis, can also be applied to SNP genotyping.

Other hybridisation methods include: PNA and LNA probes, DASH, and Qbead (for detailed descriptions see (Kim and Misra 2007).

III. Ligation (Advantage: Variety of assay formats, Disadvantage: Multiple labelled probes required)

Genotyping SNPs by the oligonucleotide ligation assay (OLA) involves discrimination by DNA ligases against mismatches at the ligation site in two adjacently hybridised oligonucleotides (Syvanen 2001). A drawback of the OLAs is that detection of each SNP requires three oligonucleotide probes, which increases the costs of these assays. Two probes are specific for each allele and bind to the template at the SNP site and the third 'common' probe binds to the template adjacent to the SNP immediately next to the allele-specific probe. If the allele-specific probe binds perfectly at the SNP site, DNA ligase will join it with the ‘common’ probe. The ligation products are then detected by either colorimetric approaches in microtitre plate wells (Samiotaki 1994) or multiplex approaches using fluorescently labelled ligation probes with different electrophoretic mobilities that can be analysed in a DNA sequencing instrument (Grossman 1994), to reveal base identity at the SNP position.

Other ligation methods include: CFET (combinatorial fluorescence energy transfer) tags and Padlock probes (for detailed descriptions see (Syvanen 2001, Kim and Misra 2007).

IV. Enzymatic Cleavage (Advantage: PCR amplification avoided, Disadvantage: requires large amount of DNA)

Enzymatic cleavage for allele discrimination is based on the ability of certain classes of enzymes to cleave DNA by recognising specific sequences and structures (Kim and Misra 2007). These enzymes can be used for discriminating between alleles when SNP sites are located in an enzyme recognition sequence and allelic differences affect recognition.

The Invader assay is based on the capability of a class of enzymes called flap-endonucleases (FENs) and engineered enzymes termed cleavases to cleave DNA molecules at specific structures (Sauer and Gut 2002). It uses three probes for genotyping a SNP, two allele-specific probes and a third common probe (invader). The invader oligonucleotide is designed with its 3’ending on the polymorphism to be tested with a mismatch at the SNP site. The allele-specific oligonucleotides are complementary to the region 5’ of the polymorphic site with an overhang at their 5’ end. After displacement of the allele-specific probes by the invader probe, FEN mediated cleavage occurs for the overhang of the allele-specific probe at the SNP site and it is detected by fluorescence. The mismatch probe remains intact and is not detected; hence the assay can be used for detection of both allelic variants (Kim and Misra 2007). The released overhang oligonucleotide can then serve as an invader for a second reaction. This two step assay provides much higher signal amplification and eliminates the need for PCR amplification. However, the technique is limited to genomic DNA or large quantities of target DNA (Syvanen 2001).

Allele detection methods

Products from the various allele-discriminating reactions described above are then analysed for allelic differences. Major detection methods include mass spectrometry, fluorescence and chemiluminescence (see Table 1).

I. Mass Spectrometry

Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF) was developed as a tool for differentiating genotypes, by comparing the mass of DNA fragments after a single ddNTP primer extension reaction (Sauer and Gut 2002). With this detection method, a mixture of many oligonucleotides can be rapidly separated and accurately analysed without the need for specific detection labels. Mass spectrometry is particularly useful as a read-out method for primer extension reactions because primers of different lengths can be used in combination with mixtures of deoxy- and dideoxynucleoside triphosphates designed to yield allele-specific primer extension products with clear differences in their molecular mass (Syvanen 2001). The assays can also be multiplexed if the primer extension products for each SNP have non-overlapping mass distributions. A limitation with MALDI-TOF is that the primer extension products must be rigorously purified before measurement to avoid background from biological material present in the sample. The GOOD assay avoids this limitation (Syvanen 2001). The PinPoint assay, MassEXTEND and SPC-SBE are other SNP genotyping methods that utilize MALDI-TOF for genotype differentiation. High levels of throughput and automation can be attained.

II. Fluorescence signal-based detection

Allele detection through the monitoring of fluorescence signals, is widely used in current genotyping technologies because its implementation is simple and detection is fast with high sensitivity (Kim and Misra 2007). A significant feature of the Human Genome Project, fluorescence detection is used for direct sequencing using capillary array electrophoresis. It uses dideoxynucleotides labelled with different fluorescent dyes in a Sanger sequencing reaction to produce a ladder of fluorescently tagged extension products (Sanger 1977, Kim and Misra 2007). The primer extension products are separated by electrophoresis and specific fluorescence signal from each extension product exposes base identity.

Fluorescence polarisation (FP) detects the increase in polarisation of fluorescence, caused by the decreased mobility of a fluorophore through a molecular mass increase (Tsuchihashi 2002). Here, the allele-specific incorporation of fluorescently labelled dideoxy-NTP can be detected as an increase of a polarised fluorescence. Essentially, any genotyping method in which the product of the allelic discrimination reaction is substantially larger or smaller than the starting fluorescent molecule can use FP as a detection method (Sobrino *et al.* 2005). Although relatively inexpensive, eliminating the need for fluorescently labelled primers, FP has the disadvantage of being a multi-step process, incorporating both PCR and post-PCR processing steps, making it unsuitable for high levels of throughput. It also has a relatively modest signal/noise ratio, as the incorporation of a fluorescently-labelled nucleotide into the primer end changes its molecular weight only by an order of magnitude, which makes reliable automated allele calling in this method a challenge (Tsuchihashi 2002).

Fluorescence resonance energy transfer (FRET) occurs when two fluorescent dyes are in close proximity to one another and the emission spectrum of one fluorophore overlaps the excitation spectrum of the other fluorophore (Sobrino *et al.* 2005). As a result, the fluorescence signal observed when the dyes are in close proximity is different from the signal emitted when they are separated from each other (Kim and Misra 2007). The proximity requirement is what makes FRET an effective detection method for a number of allelic discrimination mechanisms. Generally, any reaction that brings together or separates two dyes can use FRET as its detection method. FRET detection has, therefore, been used in primer extension and ligation reactions where the two labels are brought into close proximity to each other, as well as in hybridisation reactions and enzymatic cleavage where the neighbouring fluorescent dye pair is separated by cleavage or disruption of the stem-loop structure that holds them together (Kwok 2001). The major drawback of this method is the cost of the labelled probes required in all the genotyping approaches with FRET detection. The cleavage approaches are particularly costly as the probes are doubly labelled.

III. Chemiluminescence and Pyrosequencing

Chemiluminescence has major advantages as a detection technique, such as high signal to noise ratio, rapid detection, and feasibility for automation (Kim and Misra 2007). Pyrosequencing employs chemiluminescence-based detection for SNP genotyping using a cascade of enzymatic reactions to detect nucleotide incorporation during DNA synthesis. It involves the sequencing-by-synthesis approach in which nucleotides are added one at a time (either cyclically or sequentially) to a primer extension reaction mixture. Whenever an added nucleotide is complimentary to the template DNA strand, a pyrophosphate is released that is immediately converted to ATP by ATP sulfurylase. This ATP causes the oxidation of luciferin by luciferase, which is detected as a light signal of which the intensity is proportional to the number of incorporated nucleotides. If the added nucleotide is not complimentary to the next base in the template no light will be generated. Because the added nucleotides are known, the sequence of the template can be determined (Ronaghi 2001).

Pyrosequencing has the significant advantage over traditional Sanger sequencing in that it requires no gels or capillaries to separate extension products by size and nucleotide incorporation can be detected in real time (Hudson 2008). Although the technique provides only short read lengths of 20-30 base pairs, its capability to read flanking sequences as well as the SNP position itself, and its high specificity (i.e. non-specific binding will not generate a false signal) make it an accurate and attractive SNP genotyping method (Tsuchihashi 2002). Using an automated microtiter-based pyrosequencer instrument allows the simultaneous analysis of 96 samples within 10 minutes (the cost using standard pyrosequencing equipment is ~US69 cents per sample; http://www.pyrosequencing.com). Each round of nucleotide dispensing takes approximately 1 minute and thus offers a rapid way to determine the exact sequence of the SNPs using adjacent positions as controls (Ahmadian *et al*. 2000). Massively parallel 454 pyrosequencing, a ‘next generation’ sequencing technology, comprises roughly 300 000 sequence reads averaging 100-250 base pairs in length, offering a major increase in throughput and considered the best proven alternative to Sanger methods (costs a little over US0.01 cents per base) (Margulies 2005, Cristobal Vera 2008, Ellegren 2008, Hudson 2008). A striking feature of pyrogram readouts for SNP analysis is the clear distinction between the various genotypes; each allele combination (homozygous or heterozygous) will give a specific pattern compared with the two other variants (Ahmadian *et al.* 2000, Ronaghi 2001, Ahmadian *et al.* 2006). Thus it is relatively easy to score the allelic status by pattern recognition software. Furthermore, pyrosequencing enables determination of the phase of SNPs when they are in the vicinity if each other allowing the detection of haplotypes (Ahmadian 2006b).

The limiting factors of pyrosequencing are the template preparation required and the degree of multiplexing (Sobrino *et al.* 2005). Prior to analysis, PCR products need to be converted to single stranded template onto which a sequencing primer is annealed. As with real time PCR, the method also has the problem of limited multiplex capability.

SNP Genotyping in whales

Morin *et al.* (2007a) used a combination of primer extension and fluorescence to genotype SNPs in sperm whales. To avoid linked loci, they selected one SNP from each locus (39 SNPs were identified in 23 of the 44 loci encompassing a total length of 21 063bps) for allele-specific amplification using either Luminex X-map or Amplifluor technologies. The Luminex system requires pre-amplification of each locus using flanking PCR primers, followed by allele-specific primer extension (ASPE) with two allele-specific primers and incorporation of biotin-labelled nucleotides, with subsequent binding of ASPE products to streptavidin phycoerythrin for fluorescent detection. ASPE offers both the advantage of streamlining the SNP analysis protocol and the ability to perform multiplex SNP analysis on any mixture of allelic variants (Taylor *et al.* 2001). Here, measurement of allele-specific fluorescence is determined through flow cytometric analysis. Amplifluor technology is simpler, but doesn't allow multiplexing, and is based on fluorescence resonance energy transfer (FRET). Briefly, Amplifluor utilizes hairpin-shaped primers that function in a similar way to molecular beacons (oligonucleotide probes with complimentary bases at either end): when the primer is incorporated into a PCR product, the donor and quencher moieties are separated and donor fluorescence is thus increased. Fluorescence can be detected in real time using a quantitative PCR instrument, or after completion of the PCR using a fluorometer (Giancola *et al.* 2006). Morin and McCarthy (2007) further developed this process to allow genotyping of historical and low-quality samples by developing multiplex preamplification of all SNP loci in one PCR prior to performing individual genotyping assays as a way of improving data quality and completeness from each Amplifluor assay.

Ultimately, how one goes about identifying and genotyping SNPs will depend largely on the equipment and expertise available in the laboratory and the type of project envisaged i.e. it is quite different to perform genotypes with a limited number of SNPs on very large population samples, or a large number of SNPs on a limited number of individuals (Vignal *et al.* 2002). The number of SNPs required will also depend upon the population genetics application. Precise estimation and comparison of genetic variation among populations requires a large number of SNPs relative to microsatellites because microsatellite loci typically have many alleles, whereas two is the norm for SNP loci (Morin *et al.* 2004). It is estimated that for accurate parentage determination in natural populations, up to 100 SNPs are required (Anderson 2006) whereas with highly polymorphic microsatellite loci, as few as 3 loci have proven to be sufficient (Saino 1997). For linkage studies and estimates of population genetic parameters, approximately three times as many SNPs are needed in comparison to microsatellites (Brumfield *et al.* 2003). However, the required number of loci is difficult to assess *a priori* because each study has a different evolutionary context and thus far only few empirical studies have compared the utility of microsatellites and SNPs for population genetic studies of non-model organisms.

Another consideration when choosing SNPs for population genetic studies is the risk of ascertainment bias associated with SNP discovery which has the potential to introduce systematic bias in estimates of variation within and among populations. Informally, the more polymorphic a SNP is (by which we mean the larger the relative frequency of the less common, or minor variant) the more likely it is to be found during the SNP discovery process. Ascertainment bias is most problematic for applications that use allele frequencies to estimate population size and demographic changes, and least important for individual identification, paternity analysis, and assignment tests, where certain loci are selected intentionally (Morin *et al.* 2004). Unfortunately, most SNP discovery methods incorporate some level of ascertainment bias (Garvin and Gharrett 2007). The challenge is to reduce the bias as much as possible by first determining how a SNP should be defined (e.g. should every variable site be a SNP or should only high heterozygosity loci be chosen?) and secondly, whether all individuals in the study will be sequenced for all loci or only a panel or subset of individuals (Brumfield *et al.* 2003). These decisions on protocol will enable ascertainment bias to be assessed and potentially corrected. New SNP detection methods are also being designed to easily identify and reduce ascertainment bias as well as data analysis that considers ascertainment schemes in the analyses of SNP data. Ultimately, such bias can be avoided by using large numbers of individuals that are representative of the populations being studied in the discovery panel (Morin *et al.* 2004), however this is not always feasible. DEco-TILLING, a modification of Eco-TILLING (targeting induced local lesions in genomes) claims to discover SNPs rapidly at low cost, minimizes the number of individuals that need to be sequenced in order to genotype large numbers of individuals and reduces ascertainment bias (Garvin and Gharrett 2007). The technique is based on random double-strand cleavage at specific mispaired nucleotide sites in heteroduplexes. In terms of data analysis, Nielsen and Signorovitch (2003) have shown how appropriate modelling of realistic ascertainment schemes can be done in a variety of situations. Specifically, it is relatively easy to correct the sampling distribution for SNPs at one or two loci with combinatorial expression that models the ascertainment scheme.

Finally, the probability of linkage between SNP loci and recombination increases with the higher numbers that are often required relative to microsatellites. Recombination can influence both the interpretation and the sampling strategy of SNP variation. In humans SNP variation was found to be low in regions of low recombination, and high in regions of high recombination (Nachman 2001). This same pattern is likely to hold in many other species, both model and non-model organisms (Brumfield *et al.* 2003). Linkage between SNP loci can bias statistical estimates in models that assume independence (Vignal *et al.* 2002). Both potential complications must be kept in mind when developing SNP's for population genetic analyses, particularly when using dense sets of SNP's. It is also worth noting that unlike microsatellites that are seldom found in coding sequences and are by definition neutral, SNPs are widespread across the genome and so must be individually checked for neutrality.

It is a strongly supported view that the use of SNPs as alternative nuclear markers will be superior to microsatellite data in most, if not all, population genetic studies (Morin *et al.* 2007b) because of the potential for higher genotyping efficiency, data quality, genome-wide coverage and analytical simplicity (e.g. in modelling mutational dynamics). The strength of using numerous independent genetic markers to determine the structure and phylogeny of closely related right whale species was clearly supported by Gaines *et al*.

(2005). In this study, the inclusion of numerous independent markers helped reveal evolutionary relationships that were not discerned through the analysis of individual nuclear DNA markers. The addition of a mitochondrial marker also added to the strength of the analysis as a result of complementary resolving power. All of these factors, together with the number of SNP markers already identified in at least two whale species, would support the development and application of SNP's for humpback whale population genetic studies. SNPs have the potential to measure subtle differences in humpback population structure as well as placing historical demography and speciation studies on a common molecular framework, one that is comparable to the years of mtDNA work already undertaken.

LITERATURE CITED

Adams, M.D., Kelley, J.M., Gocayne, J.D. et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. . Science 252:1651-1656.

Ahmadian, A., M. Ehn and S. Hober. 2006. Pyrosequencing: History, biochemistry and future. Clinica Chimica Acta 363:83-94.

Ahmadian, A., B. Gharizadeh, A.C. Gustafsson, F. Sterky, P. Nyren, M. Uhlen and J. Lundeberg. 2000. Single-nucleotide polymorphism analysis by pyrosequencing. Analytical Biochemistry 280:103-110.

Ahmadian, A., Lundeburg, J., Nyren, P., Uhlen, M. and Ronaghi, M. 2006b. Analysis of the p53 tumor suppressor gene by Pyrosequencing. Bio Techniques 28:140-144.

Aitken, N., S. Smith, C. Schwarz and P.A. Morin. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Molecular Ecology 13:1423-1431.

Altshuler, D., V.J. Pollara, C.R. Cowles, W.J. Van Etten, J. Baldwin, L. Linton and E.S. Lander. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513-516.

Anderson, E.C.a.G., J.C. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567-2582.

B-Rao, C. 2001. Sample size considerations in genetic polymorphism studies. Human Heredity 52:191-200.

Baker, C.S., M.D. Vant, M.L. Dalebout, G.M. Lento, S.J. O'Brien and N. Yuhki. 2006. Diversity and duplication of DQB and DRB-like genes of the MHC in baleen whales (suborder : Mysticeti). Immunogenetics 58:283-296.

Balloux, F.a.L.-M., N. 2002. The estimation of population differentiation with microsatellite markers. Molecular Ecology 11:155-165.

Bensasson, D., D.X. Zhang, D.L. Hartl and G.M. Hewitt. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution 16:314-321.

Bensch, S., S. Akesson and D.E. Irwin. 2002. The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. Molecular Ecology 11:2359-2366.

Bierne, N., Lehnert, S.A., Bedier, E., Bonhomme, F. and Moore, S.S. 2000. Screening for intron length polymorphisms in penaeid shrimps using exon-primered intron-crossing (EPIC)-PCR. Molecular Ecology 9:233-235.

Bortolin, S., Black, M., Modi, H., Boszko, I., Kobler, D., et al. 2004. Analytical validation of the tag-it high-throughput microsphere-based univsersal array genotyping platform: application to the multiplex detection of a panel of thrombophilia-associated single nucleotide polymorphisms. Clinical Chemistry 50:2028-2036.

Brumfield, R.T., P. Beerli, D.A. Nickerson and S.V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology & Evolution 18:249-256.

Callen, D.F., A.D. Thompson, Y. Shen, H.A. Phillips, R.I. Richards, J.C. Mulley and G.R. Sutherland. 1993. Incidence and origin of "null" alleles in the $(AC)_n$ microsatellite markers. American Journal of Human Genetics 52:922-927.

Cappuccio, I., L. Pariset, P. Ajmone-Marsan, S. Dunner, O. Cortes, G. Erhardt, G. Luhken, et al. 2006. Allele frequencies and diversity parameters of 27 single nucleotide polymorphisms within and across goat breeds. Molecular Ecology Notes 6:992-997.

Cristobal Vera, J., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I. and Marden, J.H. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology 17:1636-1647.

DeBry, R.W.a.S., S. 2001. Nuclear intron sequences for phylogenetics of closely related mammals: an example using phylogeny of *Mus*. Journal of Mammalogy 82:280-288.

Dunshea, G., N.B. Barros, R.S. Wells, N.J. Gales, M.A. Hindell and S.N. Jarman. 2008. Pseudogenes and DNA-based diet analyses: a cautionary tale from a relatively well sampled predator-prey system. Bulletin of Entomological Research 98:239-248.

Ebersberger, I., D. Metzler, C. Schwarz and S. Paabo. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. American Journal of Human Genetics 70:1490-1497.

Ellegren, H. 2008. Sequencing goes 454 and takes large-scale genomics into the wild. Molecular Ecology 17:1629-1635.

Ewing, B., Hillier, L., Wendl, M.C. and Green, P. 1998a. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research 8:175-185.

Ewing, B.a.G., P. 1998b. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research 8:186-194.

Excoffier, L. and Z. Yang. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol. Biol. Evol. 16:1357-1368.

Friesen, V.L., Congdon, B.C., Kidd, M.G. and Birt, T.P. 1999. Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. Molecular Ecology 8:2147-2149.

Gagneux, P., C. Boesch and D.S. Woodruff. 1997. Microsatellite scoring errors associated with non-invasive genotyping based on nuclear DNA amplified from shed hair. Molecular Ecology 6:861-868.

Gaines, C.A., M.P. Hare, S.E. Beck and H.C. Rosenbaum. 2005. Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. Proceedings of the Royal Society B-Biological Sciences 272:533-542.

Garvin, M.R. and A.J. Gharrett. 2007. DEco-TILLING: an inexpensive method for single nucleotide polymorphism discovery that reduces ascertainment bias. Molecular Ecology Notes 7:735-746.

Giancola, S., H.I. McKhann, A. Berard, C. Camilleri, S. Durand, P. Libeau, F. Roux, et al. 2006. Utilization of the three high-throughput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants. Theoretical and Applied Genetics 112:1115-1124.

Givens, G.H., Williams, M., Morin, P.A., Hancock, B. and George, J.C. 2007. A brief and preliminary look at SNPs data for some Bering-Chukchi-Beaufort Seas bowhead whales. International Whaling Commission.

Grossman, P.D.e.a. 1994. High density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. Nucleic Acids Research 22:4527-4534.

Hare, M.P. 2001. Prospects for a nuclear gene phylogeography. Trends in Ecology & Evolution 16:700-706.

Hare, M.P.a.P., S.R. 2003. High intron sequence conservation across three mammalian orders suggest functional constraints. Mol. Biol. Evol. 20.

Harpending, H.C., M.A. Batzer, M. Gurven, L.B. Jorde, A.R. Rogers and S.T. Sherry. 1998. Genetic traces of ancient demography. Proceedings of the National Academy of Sciences of the United States of America 95:1961-1967.

Hoffman, J.I., C.W. Matson, W. Amos, T.R. Loughlin and J.W. Bickham. 2006. Deep genetic subdivision within a continuously distributed and highly vagile marine mammal, the Steller's sea lion (*Eumetopias jubatus*). Molecular Ecology 15:2821-2832.

Hudson, M.E. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. Molecular Ecology Reesources 8:3-17.

Kennedy, G.C., Matsuzaki, H., Dong, S., Lui, W.M., Huang, J., et al. 2003. Large-scale genotyping of complex DNA. Nat. Biotechnol. 21:1233-1237.

Kim, S. and A. Misra. 2007. SNP genotyping: Technologies and biomedical applications. Annual Review of Biomedical Engineering 9:289-320.

Krause, J., Dear, P.H., Pollack, J.L. et al. 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature 439:724-727.

Kwok, P. 2001. Methods for genotyping single nucleotide polymorphisms. Annu. Rev. Genomics Hum. Genet. 2:235-258.

LaHood, E.S., P. Moran, J. Olsen, W.S. Grant and L.K. Park. 2002. Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. Molecular Ecology Notes 2:187-190.

Lessa, E. 1992. Rapid surveying of DNA sequence variation in natural populations. Molecular Biology & Evolution 9:323-330.

Lyons, L.A., Laughlin, T.F., Copeland, N.G. et al. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nature Genetics 15:47-56.

Margulies, M., Egholm, M., Altman, W.E., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Marth, G.T., Korf, I., Yandell, M.D., Yeh, R.T., Gu, Z., Zakeri, H., Stitziel, N.O., Hillier, L., Kwok, P.Y. and Gish, W.R. 1999. A general approach to single nucleotide polymorphism discovery. Nature Genetics 23:452-456.

McLenachan, P.A., K. Stockler, R.C. Winkworth, K. McBreen, S. Zauner and P.J. Lockhart. 2000. Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. Molecular Ecology 9:1899-1903.

Morin, P.A., N.C. Aitken, N. Rubio-Cisneros, A.E. Dizon and S. Mesnick. 2007a. Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*). Molecular Ecology Notes 7:626-630.

Morin, P.A., K.E. Chambers, C. Boesch and L. Vigilant. 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Molecular Ecology 10:1835-1844.

Morin, P.A., B.L. Hancock and J.C. George. 2007b. Development and application of single nucleotide polymorphisms (SNPs) for bowhead whale population structure analysis. Paper SC/59/BRG8 presented to the IWC Scientific Committee, 2007 (unpublished). 7pp.

Morin, P.A., G. Luikart and R.K. Wayne. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution 19:208-216.

Morin, P.A. and M. McCarthy. 2007. Highly accurate SNP genotyping from historical and low-quality samples. Molecular Ecology Notes 7:937-946.

Nachman, M.W. 2001. Single nucleotide polymorphisms and recombination rate in humans. Trends In Genetics 17:481-485.

Nachman, M.W. and S.L. Crowell. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics 156:297-304.

Navidi, W., N. Arnheim and M.S. Waterman. 1992. A multiple tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. American Journal of Human Genetics 50:347-359.

Nickerson, D.A., Tobe, V.O. and Taylor, S.L. 1997. Polyphred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Research 25:2745-2751.

Nicod, J.C. and C.R. Largiader. 2003. SNPs by AFLP (SBA): a rapid SNP isolation strategy for non-model organisms. Nucleic Acids Research 31.

Nielsen, R. and J. Signorovitch. 2003. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. Theoretical Population Biology 63:245-255.

Oakley, T.H.a.P., R.B. 1999. Phylogeny of salmonine fishes based on growth hormone introns: Atlantic (*Salmo*) and Pacific (*Oncorhynchus*) salmon are not sister taxa. Molecular Phylogenetics and Evolution 11:381-393.

Palumbi, S.R. and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtdna of humpback whales. Molecular Biology & Evolution 11:426-435.

Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L. and Syvanen, A.C. 2000. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Research 10:1031-1042.

Picoult-Newberg, L., T.E. Ideker, M.G. Pohl, S.L. Taylor, M.A. Donaldson, D.A. Nickerson and M. Boyce-Jacino. 1999. Mining SNPs from EST databases. Genome Research 9:167-174.

Primmer, C.R., T. Borge, J. Lindell and G.P. Saetre. 2002. Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. Molecular Ecology 11:603-612.

Prychitko, T.M.a.M., W.S. 1997. The utility of DNA sequences of an intron from the beta-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). Molecular Phylogenetics and Evolution 8:193-204.

Rompler, H., Dear, P.H., Krause, J. et al. 2006a. Multiplex amplification of ancient DNA. Nature Protocols 1:720-728.

Rompler, H., Rohland, N., Lalueza-Fox, C. et al. 2006b. Nuclear gene indicates coat-colour polymorphism in mammoths. Science 313:62-62.

Ronaghi, M. 2001. Pyrosequencing sheds light on DNA sequencing. Genome Research 11:3-11.

Ryynanen, H.J. and C.R. Primmer. 2006. Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. Bmc Genomics 7.

Saino, N., Primmer, C.R., Ellegren, H. and Moller, A.P. 1997. An experimental study of paternity and tail ornamentation in the barn swallow (*Hirundo rustica*). Evolution 51:562-570.

Samiotaki, M., Kwiatkowski, M., Parik, J. and Landegren, U. 1994. Dual-colour detection of DNA sequence variants by ligase-mediated analysis. Genomics 20.

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 96:5463-5467.

Sauer, S. and I.G. Gut. 2002. Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 782:73-87.

Schlotterer, C. 2004. The evolution of molecular markers - just a matter of fashion? Nature Reviews Genetics 5:63-69.

Seddon, J.M., H.G. Parker, E.A. Ostrander and H. Ellegren. 2005. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Molecular Ecology 14:503-511.

Silva, J.C. and A.S. Kondrashov. 2002. Patterns in spontaneous mutation revealed by human-baboon sequence comparison. Trends In Genetics 18:544-547.

Sobrino, B., M. Brion and A. Carracedo. 2005. SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Science International 154:181-194.

Sokolov, B.P. 1990. Primer extension technique for the detection of single nucleotide in genomic DNA. Nucleic Acids Research 18:3671.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in Ecology & Evolution 15:199-206.

Syvanen, A.C. 2001. Accessing genetic variation: Genotyping single nucleotide polymorphisms. Nature Reviews Genetics 2:930-942.

Taberlet, P., S. Griffin, B. Goosens, S. Questiau, V. Manceau, N. Escaravage, L.P. Waits, et al. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Research 24:3189-3194.

Taylor, J.D., D. Briley, Q. Nguyen, K. Long, M.A. Iannone, M.S. Li, F. Ye, et al. 2001. Flow cytometric platform for high-throughput single nucleotide polymorphism analysis. Biotechniques 30:661-+.

Tsuchihashi, Z.a.D., N.C. 2002. Progress in high throughput SNP genotyping methods. The Pharmacogenomics Journal 2:103-110.

Ugozzoli, L., Wahlqvist, J.M., Ehsani, A., Kaplan, B.E. and Wallace, R.B. 1992. Detection of specific alleles by using allele-specific primer extension followed by capture on solid support. Genet. Anal. Tech. Appl. 9:107-112.

Vignal, A., D. Milan, M. SanCristobal and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34:275-305.

Wolford, J.K., Blunt, D., Ballecer, C. and Prochazka, M. 2000. High throughput SNP detection by using DNA pooling and denaturing high performance liquid chromatography (DHPLC). Human Genetics 107:483-487.

Table 1. Allele detection methods that can be used for each allele discrimination reaction to genotype SNPs.

Detection method	Mass spectrometry	Fluorescence			Chemiluminescence
Discrimination method		Electrophoresis	FRET	FP	
Primer Extension	X	X	X	X	X
Hybridisation			X	X	
Ligation		X	X		
Enzymatic Cleavage	X		X	X	