**FACULDADE DE BIOCIÊNCIAS**

**PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA**

# ESTRUTURA POPULACIONAL E HISTÓRIA DEMOGRÁFICA DAS POPULAÇÕES DE BALEIAS JUBARTE (*Megaptera novaeangliae*) DA AMÉRICA DO SUL

**Ana Lúcia Cypriano de Souza**

**TESE DE DOUTORADO**

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL**

**Av. Ipiranga 6681 - Caixa Postal 1429**

**Fone: (051) 3320-3500 - Fax: (051) 3339-1564**

**CEP 90619-900 Porto Alegre - RS**

**Brasil**

**2013**

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA

# ESTRUTURA POPULACIONAL E HISTÓRIA DEMOGRÁFICA DAS POPULAÇÕES DE BALEIAS JUBARTE (*Megaptera novaeangliae*) DA AMÉRICA DO SUL

**Ana Lúcia Cypriano de Souza**
**Orientador: Dr. Sandro Luis Bonatto**

**TESE DE DOUTORADO**
**PORTO ALEGRE - RS - BRASIL**

**2013**

**Sumário**

Às baleias jubarte que são seres
fascinantes e foram minha fonte de inspiração.

**Agradecimentos**

Para escrever essa tese eu precisei muito mais do que a ajuda de pesquisadores e colegas. Esse estudo é o resultado da interação de várias pessoas que contribuíram para a minha jornada, não só durante os quatro anos e meio de doutorado, mas durante os nove anos trabalhando com as baleias jubarte e os oito anos dedicados a genética da conservação dessa espécie.

Primeiramente agradeço à minha família que à distância me apoiou durante todos esses anos, em especial à minha mãe Yolanda pelo amor, amizade, por me incentivar e acreditar na realização deste meu sonho. Ao meu pai Wagner pelo amor, preocupação, compreensão e apoio. Ao meu irmão Leonardo (*In memorian*) que me incentivou desde o início e foi meu guia ("pastor" de baleias) durante todos esses anos. Às minhas irmãs, Andréa e Luciana, pelo amor, amizade e apoio. Aos meus sobrinhos, Mariana, João Pedro, Luiza e Tiago por serem tão especiais e pelo carinho. Agradeço muito por ter uma família como essa. Vocês representam tudo para mim. Amo muito!

Ao meu amor Luciano, pela paciência e compreensão durante os seis anos que estamos juntos, mas morando em cidades distantes. Obrigada por sempre me escutar, pelo incentivo, carinho e amor, o que me dá muita força. Amo te!

Às baleias jubate (jujubas) por serem seres tão fascinantes, que despertaram meu interesse em saber sobre sua biologia, além da vontade de contribuir de alguma forma para a sua preservação. Nunca foi difícil ser inspirada por vocês.

Em 2004 começou a minha trajetória com as baleias jubarte. No Instituto Baleia Jubarte iniciei como estagiária do Projeto Sotalia (boto cinza), aonde tive a oportunidade de conhecer essa espécie e de fazer amigos, em especial à Patrícia Bueno pela amizade. Agradeço ao orientador Marcos Rossi-Santos pela oportunidade e por permitir a minha ida para o Projeto Baleia Jubarte, para finalmente estagiar com as jubartes. Ao Milton Marcondes por me indicar para o estágio. À Luiza Pacheco pela oportunidade do estágio. À Renata Santoro Sousa Lima pela maravilhosa chance de escutar as jubartes pela primeira vez....inesquecível. À Grazyela Fiuza pelos dias divertidos no Tomara. À Claudia Di Petta e Tati Bisi pela amizade e incentivo. Em especial à Maria Emília Morete (Mia) pelos maravilhosos dias na ilha de Santa Barbára fazendo o que mais amo...observando as "jujubas". Obrigada pelas conversas, pelo incentivo e pelos conselhos. Aos estagiários e amigos, Chris Del Vechio, Lucas Pilli, Márcia Araújo, Patrícia Golodne, em especial à amigas e irmãs Luciana Garrido e Fernanda Nogueira por todos esses anos de amizade.

Agradeço especialmente à Márcia Engel por ter me dado a oportunidade de realizar este trabalho, por ter acreditado na minha capacidade, pela confiança, incentivo e suporte ao trabalho durantes todos esses anos. Obrigada mais uma vez!

À toda equipe do Instituto Baleia Jubartes, principalmente aos pesquisadores e aos estagiários que coletaram as amostras em campo.

Finalmente em março de 2005 chego em Porto Alegre para iniciar um trabalho que foi um grande desafio para mim, já que não tinha experiência no campo da ecologia molecular. Durante todos esses anos de aprendizado, eu fui inspirada e motivada a gostar cada vez mais dessa área, na qual utilizamos ferramentas genéticas para aprender sobre espécies incríveis, como a baleia jubarte, e podemos ajudar nas decisões de manejo. Várias pessoas contribuíram direta ou indiretamente na realização desse trabalho gratificante.

Agradeço em especial ao meu orientador Sandro L. Bonatto, pela confiança e orientação desde o meu mestrado, me deixando mais autônoma e confiante. Mas principalmente pela paciência já que cheguei ao laboratório sem experiência em genética, e até hoje encontro dificuldade nas análises de alguns programas, com os quais você me ajudou muito. Agradeço também pelo seu apoio e compreensão nas etapas mais difíceis tanto no mestrado quanto no doutorado. Obrigada por tudo!

Agradeço à todos genômicos (ex e atuais) por fazerem do laboratório um ambiente de alegria e descontração, pelo incentivo e ajuda durante a realização deste trabalho. Várias pessoas passaram pelo Genoma durante esses anos e compartilharam vários momentos: Alice Ribeiro, Ana Liedke, André Schnorr (*In memorian*), Fê Britto, Felipe Grazziotin, Gabriel Macedo, Karla Yotoko, Kiko (das tartarugas), Nelson Fagundes, Paulinho Prates, Paulinho Chaves, Ricardo Kanitz, Roberta Graboski, Thomas Xavier. A maioria dos colegas de trabalho se tornaram grandes amigos. Especialmente às meninas do "Encontrinho da fofoca": Alexsandra Schneider, Anelise Hahn, Carla Pires, Cris Trinca, Déa Thomas, Flávia Tirelli, Fê Pedone, Gabi Fernández, Laura Moretti, Anelisie Santos, Marina Favarini, Mirian Tsuchiya, Rê Bornholdt, Taia Haag, Talita Pereira. Obrigada por todos os momentos, nos vários encontrinhos, nas festinhas genômicas, e no dia a dia no lab. Obrigada pelas conversas, pelas risadas e pelo incentivo nas horas mais difíceis. Vocês são MARA e vou sentir muita saudade. Espero pelo menos uma vez por ano nos encontrarmos...EVERYBODY TOGETHER em um encontrinho da fofoca.

À Cladinara Sarturi e Priscilla Zamberlan que fizeram do laboratório um local sempre organizado, facilitando assim o trabalho de todos e pela disponibilidade para responder minhas dúvidas mais frequentes.

À Larissa Rosa de Oliveira e à Raquel Almeida que me ensinaram as primeiras práticas laboratoriais e estavam sempre dispostas a ajudar.

Ao Carlos André Lima Rosa que me apresentou ao mundo dos microssatélites, e ao Manoel Fontoura Rodrigues, que com toda paciência, me ensinou a fazer PCR e sempre que podia tirava minhas dúvidas.

Ao Tiago Ferraz (meu fio) que iniciou no lab como bolsista de iniciação trabalhando junto comigo e, me ajudou muito nesses últimos 3 anos. Obrigada meu fio!

Às minhas queridas amigas Fernanda Silveira, Luciana Garrido e Patrícia Bueno por terem me recebido tão bem em Porto Alegre, me apresentado à cidade e me deixarem fazer parte de suas vidas, o que me deu força para permanecer longe da família. Obrigada pela grande amizade e pela compreensão da minha ausência nesses dois últimos anos. Em especial à Tia Sílvia e à Fê, que se tornaram minha família gaúcha e me deram amor, amizade, incentivo e apoio. Obrigado por tudo que vocês fizeram por mim. Amo vocês!

Às minhas amigas Taia, Ale e Cris, que dividiram comigo vários momentos, de alegria e tristeza, sempre me escutando e apoiando nas horas mais difíceis. Obrigada pela amizade, pelo companheirismo, pelas risadas e por toda ajuda e incentivo. Amooo!!!

Finalmente, no último ano do doutorado tive a maravilhosa oportunidade de fazer meu doutorado sanduíche com o Dr. Charles Scott Baker. Desde que entrei para o mundo da genética da conservação de cetáceos, essa chance foi a realização de um sonho para mim. Eu já tinha uma grande admiração por você como pesquisador, que só aumentou mais com a convivência durante os seis meses, e agora admiro também como excelente pessoa. Sua dedicação à cada um dos seus alunos de pós-graduação é surpreendente, e desde que cheguei me senti bem recebida e fazendo parte do grupo. Agradeço pela oportunidade ao me aceitar como aluna visitante do seu laboratório, pela orientação, pelos ensinamentos e discussões na reuniões semanais do laboratório, aonde aprendi muito em pouco tempo. Obrigada principalmente pelo respeito e pela confiança, que instigou em mim competências e valores que irão me ajudar a tornar uma boa cientista. Sou muito grata também pelo carinho de você e da Anjanette Baker que me receberam muito bem e fizeram com que eu me sentisse em casa durante minha estadia em Newport, no Oregon. Com certeza foi uma das melhores experiências que tive em minha vida.

À todos do *Cetacean Conservation and Genomics Laboratory*, por me receberem no laboratório, pela assistência, pelas discussões e ensinamentos durante o doutorado sanduíche. Foi uma ótima experiência trabalhar em um laboratório referência na genética de cetáceos em nível internacional. Com certeza foi uma grande motivação. À Angie Sremba pela ajuda nas

análises dos haplótipos de DNA mitocondrial. À Beth Slikas por me ajudar e ensinar as técnicas moleculares do sequenciador 454 Junior. À Alana Alexander que sempre que podia tirava minhas dúvidas. À Debbie Steel por me ajudar na calibração dos microssatélites. À Becca Hamner por me ensinar as análises do GeneClass, pelas conversas, e pela parceria em *hikings* e *sushis*. Em especial à Renee Albertson pela parceria da *bike* nos finais de semana, *hikings*, *sushis* e excelentes *beers*. Obrigada pela amizade, pelas ótimas conversas e risadas. À Maria Jose Perez, obrigada por compartilhar os sentimentos latinos, pela parceria e pelas várias conversas no lab ou no alojamento.

À todos os funcionários e pesquisadores do *Hatfield Marine Science Center*, da *Oregon State University*. Em especial a Melody Pfister pela organização dos alojamentos e pela disponibilidade em ajudar. Ao Manny pelas conversas nas noites de trabalho no laboratório. Ao John Chapman pelas conversas. Ao Markus Horning e sua esposa Lisa Mulcahy por confiarem duas vezes sua casa e seu gato, Toby, aos meus cuidados.

À Lidia, amiga brasileira que me ajudou a comprar meu carro nos *States*: o Volvão.

Aos grandes amigos feitos na América: Niki Diogou, Eric Brausser, Matt Yergey e Katlyn Taylor. Obrigada pelas parcerias, pelas conversas e risadas. Vocês são inesquecíveis!

À Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) pela infra-estrutura. Aos professores do Programa de Pós-Graduação em Zoologia por contribuírem para minha formação. Aos coordenadores Júlio César Bicca-Marques e Roberto Reis pela disponibilidade e eficiência na coordenação. Às secretárias Luiza e Luana que sempre estiveram presentes e foram prestativas.

E finalmente à CAPES pela bolsa de doutorado concedida, e pela bolsa PDSE concedida para a realização do meu doutorado sanduíche no exterior.

**Resumo**

A caça comercial baleeira durante o século XX reduziu significativamente a maioria das populações de baleias jubarte (*Megaptera novaeangliae*). Sete estoques reprodutivos (A-G) são reconhecidos pela Comissão Internacional Baleeira (CIB) no Hemisfério Sul. As baleias jubarte do estoque reprodutivo A são distribuídas ao longo da costa brasileira (principalmente entre 5° e 23° S), no Oceano Atlântico Sul Ocidental, enquanto as jubartes do estoque G ocorrem da costa do Peru (6° S) até a Costa Rica (12° N), no Oceano Pacífico Oriental. Apesar de estudos anteriores terem fornecido importantes informações sobre ambos estoques reprodutivos Sul Americanos, o grau de conectividade e de diferenciação entre essas populações precisa ser melhor investigado. Deste modo, o manuscrito 2 desta tese representa a primeira análise de diferenciação genética e nível de fluxo gênico entre essas populações, usando sequências de DNA mitocondrial e 16 locos de microssatélites. Nossos resultados revelaram uma significante diferenciação entre os estoques A e G em ambos marcadores moleculares (DNAmt e microssatélites), especialmente através da análise bayesiana que identificou duas populações mesmo sem informação dos locais de amostragem. No entanto, os testes de *assignment* indicaram um intercâmbio de indivíduos entre essas populações, mas com um fluxo gênico baixo o suficiente permitindo a independência demográfica desses dois estoques. Nossos dados separados por sexo apresentaram uma diferenciação genética significativa entre as fêmeas do Brasil e da Colômbia, e entre os machos do Brasil e da Península Antártica, sugerindo maior fidelidade das fêmeas às áreas de reprodução e dos machos às áreas de alimentação. Apesar disso, estudos recentes têm demonstrado fêmeas realizando longos movimentos entre áreas de reprodução. Portanto, um esforço de amostragem principalmente na chegada e na saída das baleias migrando para essas áreas de reprodução é necessário para melhor compreender o padrão migratório das jubartes dessas populações. Embora a população de baleias jubarte do Brasil tem demonstrado sinais de recuperação após sofrer uma redução estimada a 2% de seu tamanho histórico até meados de 1950, nenhum estudo genético tem fornecido estimativas de tamanho efetivo e de censo, atual e histórico, para essa população. Para uma melhor compreensão do impacto da caça nessa população, sua história demográfica foi investigada utilizando diferentes marcadores moleculares e diferentes métodos (manuscritos 1 e 3). No primeiro manuscrito dez locos de microssatélites foram usados para estimar pela primeira vez o tamanho atual dessa população. No manuscrito 3 foi usado pela primeira vez a tecnologia de sequenciamento em larga escala para sequenciar múltiplos locos nucleares em 24 amostras de jubartes brasileiras. Apesar da análise de computação Bayesiana aproximada suportar um cenário de população constante sobre os cenários de mudança do tamanho da população durante a caça comercial, nossas estimativas de tamanho atual em diferentes períodos de tempo demonstraram uma flutuação do tamanho da população durante esse período (~ 2 a 4 gerações atrás). Além disso, os dados de múltiplos locos indicaram um declínio de população mais recente causado pela exploração antropogênica nos últimos 200 anos. Nossa estimativa de abundância histórica (~ 148.000 indivíduos) indica que a população de jubartes do estoque A foi muito maior do que aquele estimado (~ 24.700 indivíduos) pelos registros da caça. Finalmente, os dados dos locos nucleares também indicaram que a população estava declinando desde seu tamanho máximo atingido a cerca de 30 mil anos atrás, possivelmente relacionada com as mudanças climáticas causadas pelos ciclos de glaciação/interglaciação. Esses resultados sugerem que a população do Oceano Atlântico Sul Ocidental era maior antes do início da caça, o que deve explicar a discrepância encontrada entre as estimativas de tamanho da população, genéticas e baseadas em dados da caça.

**Abstract**

**Population structure and demographic history of the humpback whale (*Megaptera novaeangliae*) populations from South America**

Commercial whaling mainly during the 20th century reduced most populations of humpback whale (*Megaptera novaeangliae*). Seven breeding stocks (A-G) are recognized by the International Whaling Commission (IWC) in the Southern Hemisphere. Humpback whales from Breeding stock A (BSA) are distributed along the Brazilian coast (mainly between 5° and 23° S), in the Southwestern Atlantic, while the humpbacks from breeding stock G (BSG) occur from Peru (6° S) to Costa Rica (12° N) coast, in the eastern Pacific Ocean. Despite previous studies have provided important information about both South America breeding grounds, the degree of connectivity and differentiation between these populations needs to be better investigated. Therefore, the manuscript 2 of this thesis represents the first analysis of the genetic differentiation and level of gene flow between these populations, using mitochondrial DNA sequences and 16 microsatellite loci. Our results showed a significant differentiation between Breeding Stocks A and G, at both molecular markers (mtDNA and microsatellites), in specially through the Bayesian clustering analysis that identified two populations even without sampling location information. However, the assignment tests have indicated an exchange of individuals between these populations, but with a gene flow low enough to allowing the demographic independence of these two stocks. Our data segregated by gender showed a significant differentiation between females from Brazil and Colombia, and between males from Brazil and Antarctic Peninsula, suggesting higher fidelity of females to the breeding areas and of males to the feeding areas. Nevertheless, recent studies have shown females undertaking long movements between breeding grounds. Thus, a sampling effort mainly on arrival and departure of the whales migrating for these areas is needed for a better understanding of the migratory pattern of the humpbacks of these populations. Although the Brazilian humpback whale population has shown signs of recovery after suffering a reduction estimated to 2% of its historical size in the late 1950s, no genetic study has provided estimates of effective and census size, contemporary and historical, for this population. For a better understanding of the whaling impact on this population, its demographic history was investigated using different molecular markers and methods (manuscripts 1 and 3). In the first manuscript ten microsatellite loci were used to estimate for the first time its contemporary population size. In the manuscript 3 was used for the first time the high throughput sequencing technology to sequence multiple nuclear loci at 24 Brazilian humpback samples. Despite the approximate Bayesian computation analysis has supported a scenario of constant $N_e$ over size changes scenarios during the whaling period; our estimates of contemporary size at different time frames have detected a fluctuation of the population size during this period (~ 2 to 4 generations ago). Moreover, multiple sequence loci data have indicated a most recent bottleneck caused by anthropogenic population depletion over past 200 years. Our estimate of historical abundance (~ 148,000 individuals) indicates that BSA humpback population was much larger than that estimated (24,700 individuals) by whaling catch records. Finally, an extended Bayesian skyline plot of the nuclear loci indicated that the population was declining ever since a size peak around 30,000 years ago, which may be associated with the climate changes caused by glacial/interglacial cycles. These results suggest that Southwestern Atlantic humpback population was higher before the onset of the whaling period, which may explain the discrepancy found between previous genetic and catch record population size estimates at this period.

**Apresentação**

Esta tese é apresentada sob forma de artigos científicos e organizada em cinco capítulos, sendo que o primeiro capítulo (Capítulo 1) é uma introdução sobre a baleia jubarte, espécie alvo deste estudo, destacando principalmente as populações de jubartes da costa leste e oeste da América do Sul (estoques reprodutivos A e G, respectivamente) e seu atual status de conservação; e o último capítulo (Capítulo 5) se refere as conclusões gerais dos resultados obtidos dos três manuscritos científicos. Todos os outros três capítulos são escritos no formato de artigo com figuras, tabelas e material suplementar incluídos no final de cada capítulo de acordo com as normas da revista *Journal of Heredity*, à qual será submetido o primeiro manuscrito. Segue abaixo uma breve descrição do contexto de cada artigo (Capítulos 2 à 4):

**Capítulo 2** - Genetic diversity and effective and census population size of the humpback whales (*Megaptera novaeangliae*) wintering off Brazil (Breeding Stock A)

A maioria das populações de baleias jubarte no Hemisfério Sul foram reduzidas pela caça comercial baleeira durante o século XX. A população de baleias jubarte do Oceano Atlântico Sul Ocidental, conhecida como estoque reprodutivo A, tem demonstrado sinais de recuperação após sofrer uma redução estimada (baseada em dados da caça) a 2% de seu tamanho histórico. Prévios estudos moleculares demonstraram uma alta diversidade genética tanto nuclear quanto mitocondrial para as jubartes do estoque A, e nenhum sinal de *bottleneck* genético foi detectado para essa população. Esses dados são consistentes com a hipótese de que a caça comercial não durou gerações suficientes ou não reduziu suficientemente o tamanho da população para reduzir significantemente a variabilidade genética dessa população. No entanto, não existem estimativas baseadas em dados genéticos do tamanho atual e histórico para essa população. Portanto, este artigo teve como propósito investigar a história demográfica da população de jubartes do estoque A, através de estimativas do tamanho efetivo e de censo atual e histórico dessa população. Para este estudo foram utilizados dez locos de microssatélites. Esse artigo será submetido para a revista *Journal of Heredity*.

**Capítulo 3** - Genetic differentiation between humpback whales (*Megaptera novaeangliae*) from Atlantic and Pacific breeding grounds of South America

As populações de baleias jubarte permanecem durante o verão nas áreas de alimentação nas altas latitudes, e migram para as áreas de reprodução nas baixas latitudes, onde permanecem durante o inverno. Essas populações formam estoques aparentemente

distintos. Com base na distribuíção das jubartes nas áreas de reprodução no Hemisfério Sul, a Comissão Internacional Baleeira (CIB) reconhece sete estoques reprodutivos, de A a G. Recentemente, estudos genéticos de estrutura populacional usando DNA mitocondrial (DNAmt) foram realizados para as jubartes das áreas de reprodução nos oceanos Pacífico Sul, Atlântico Sul e Índico. Apesar da ausência de barreiras geográficas, a diferenciação genética (baseada no DNAmt) significativa entre essas áreas de reprodução sugere que a dispersão das fêmeas é limitada. No entanto, essa diferenciação é baixa e um grau de fluxo gênico ocorre entre esses estoques, sugerindo uma dispersão maior dos machos. Como os estudos anteriores usaram apenas dados da linhagem maternal (DNAmt), um estudo comparativo através de locos de microssatélites (biparentais) entre essas áreas é necessário para dar ou não suporte a distinção desses estoques reprodutivos. Embora os estoques reprodutivos A e G, localizados respectivamente na costa leste e oeste da América do Sul, tenham apresentado um baixo compartilhamento de haplótipos de DNAmt, o primeiro registro de intercâmbio entre essas áreas foi evidenciado através da foto-identificação de uma fêmea avistada no Equador em 1996 e reavistada no Banco dos Abrolhos em 1998. Deste modo, este artigo teve como objetivo avaliar a estrutura populacional e o nível de fluxo gênico entre as baleias jubarte dos estoques A e G. Além disso, foi também investigado a existência de um padrão de dispersão influenciado pelo macho ou pela fêmea entre essas populações. Baseado nos dados anteriores, espera-se uma fraca estruturação genética nuclear entre as jubartes desses estoques. Para este estudo foram utilizados os dois marcadores moleculares: sequências da região controle do DNA mitocondrial (465 pares de base) e 16 locos de microssatélites. O artigo será submetido para a revista *Marine Ecology Progress Series*.

**Capítulo 4** - The demographic history of the Southwestern Atlantic humpback whales (*Megaptera novaeangliae*) inferred from multiple nuclear loci suggests that the population was declining before whaling

Durante a caça comercial no século XX mais de 200.000 baleias jubarte foram caçadas no Hemisfério Sul, reduzindo algumas populações a pequenas porcentagens do seu tamanho original (pré-exploração). A Comissão Internacional Baleeira (CIB) recomenda a necessidade da estimativa de abundância das populações de baleias antes do início da caça e a reconstrução da trajetória histórica para avaliar o impacto da caça e da recuperação (se houve) dessas populações, auxiliando assim no estabelecimento dos planos de manejo. A maioria das estimativas de abundância das populações de baleias pré-exploração são baseadas em dados da caça, e nas estimativas de abundância atual, levando em conta as taxas de natalidade e

mortalidade. No entanto, essas estimativas de abundância histórica devem ser subestimadass devido à perda dos dados ou aos registros inacurados. Portanto, o propósito deste estudo foi inferir a história demográfica da população de baleias jubarte do Oceano Atlântico Sul Ocidental, estimando o tamanho efetivo e de censo histórico da população, e avaliando as flutuações populacionais ao longo do tempo. Neste artigo foi utilizada pela primeira vez a tecnologia do sequenciamento de segunda geração (454 GS Junior) para sequenciar múltiplos locos nucleares de 24 amostras de baleia jubarte, coletadas na costa do Brasil. Este artigo será submetido para a revista *PNAS*.

## Capítulo 1: Introdução Geral

A baleia jubarte (*Megaptera novaeangliae* Borowski, 1781) é provavelmente a espécie mais estudada de todas as espécies de misticetos. A espécie é encontrada em todos os oceanos (exceto o Oceano Ártico) do mundo, sendo distribuída dentro de três populações oceânicas principais: a do Pacífico Norte, a do Atlântico Norte e a do Hemisfério Sul, as quais são parcialmente separadas por barreiras físicas ou temporais (Baker *et al*. 1994, 1998, Clapham & Mead 1999). No entanto, sua distribuição geográfica é sazonalmente dependente devido à história de vida migratória da espécie, na qual subpopulações dentro de cada bacia oceânica realizam migrações de longa distância entre as altas latitudes, onde se alimentam durante o verão, e as baixas latitudes, onde acasalam e têm seus filhotes durante os meses de inverno (Mackintosh 1965, Dawbin 1966).

A maioria das áreas de alimentação e reprodução das baleias jubarte são bem conhecidas e as populações mostram diferentes níveis de fidelidade à essas áreas (Figura 1). No Hemisfério Norte as populações apresentam alta fidelidade maternal às suas áreas de alimentação, e migram para uma única área de reprodução no oceano Atlântico Norte, e para áreas de reprodução em comum no oceano Pacífico Norte (Baker *et al*. 1986, 1994, 1998, Palsboll *et al*. 1995, 1997, Calambokidis *et al*. 2001, 2008, Stevick *et al*. 2006a). No Atlântico Norte as jubartes segregam em quatro principais áreas de alimentação: Golfo do Maine, leste do Canadá, oeste da Groelândia e no leste deste oceano (Noruega, Islândia e oeste da Groelândia), sendo que as baleias dessas áreas migram para uma única área de reprodução nas Índias Ocidentais (República Dominicana) (Mattila *et al*. 1994, Palsboll *et al*. 1997, Stevick *et al*. 2006a). No entanto, recentes avistagens de baleias jubarte nas ilhas de Cabo Verde no leste do Atlântico Norte sugerem uma reocupação de uma antiga área de reprodução neste oceano (Mackintosh 1946, Reeves *et al*. 2002, Jann *et al*. 2003). No Pacífico Norte, enquanto as baleias jubarte que se alimentam ao longo da costa dos estados de Washington, Oregon e Califórnia migram principalmente para as áreas de reprodução na costa do México, e em menor número para a costa da Costa Rica (Acevedo-Gutiérrez & Smultea 1995, Calambokidis *et al*. 1996, Urbán *et al*. 2000); as jubartes das áreas de alimentação no Alasca migram principalmente para as águas em torno do Havaí (Baker *et al*. 1986, Calambokidis *et al*. 2001). No entanto, no oeste do oceano Pacífico Norte a localização das áreas de alimentação das jubartes que se reproduzem em torno das ilhas de Okinawa e Ogasawara é desconhecida (Calambokidis *et al*. 2001).

Ao contrário das segregadas áreas de alimentação das baleias jubarte no Hemisfério Norte, as áreas de forrageio das jubartes no Hemisfério Sul são distribuídas ao redor de uma larga área circumpolar do Oceano Sul sem barreira continental para impedir a dispersão, aumentando assim a probabilidade de migração entre as populações (Hoelzel 1998). Com base nos registros da caça às baleias e nos dados biológicos, essas áreas de alimentação foram divididas em seis extensas zonas de I a VI pela Comissão Internacional Baleeira (CIB ou IWC – *International Whaling Commission*), que funcionavam como unidades controle na divisão das cotas de caça comercial baleeira na Antártida (Donovan 1991). Recentemente, com base nas distribuições das jubartes em baixas latitudes durante o inverno, a CBI também propôs a divisão de sete áreas de reprodução (de A à G) no Hemisfério Sul (IWC 2005). Além disso, a população de baleias jubarte que permanece o ano todo no Mar da Arábia tem sido reconhecida pela CBI como estoque X (Mikhalev 1997, IWC 2005).

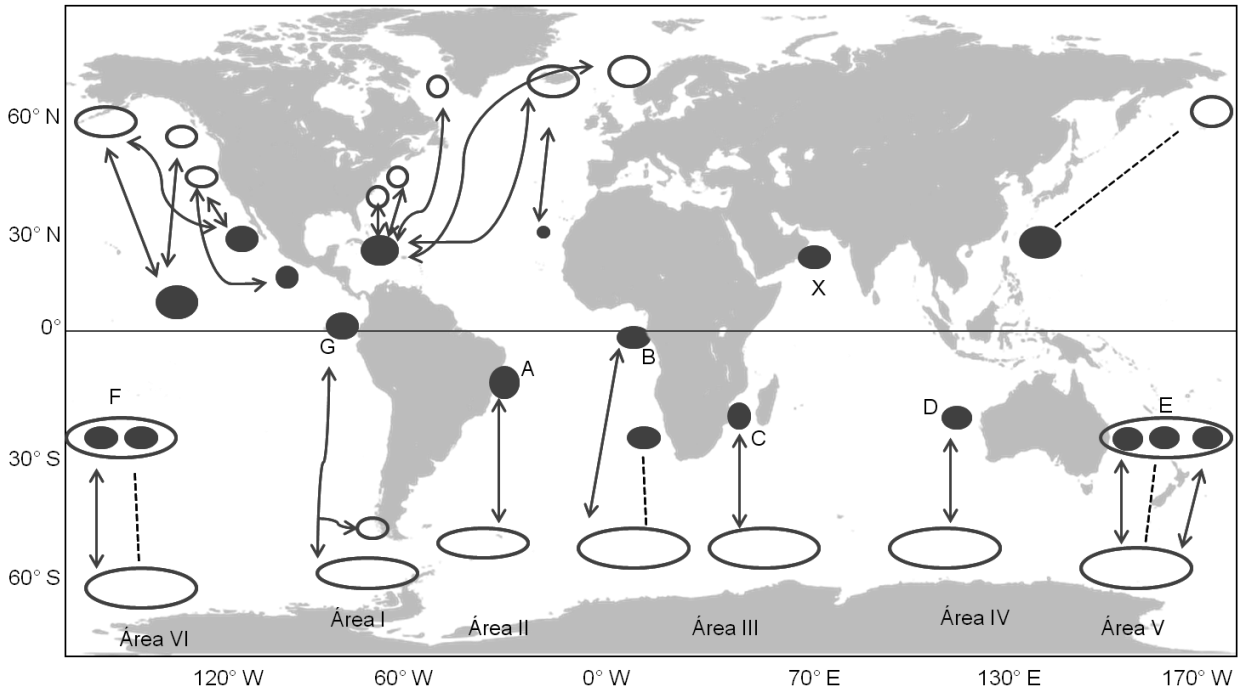



Figura 1: Distribuição das populações de baleias jubarte nas áreas de alimentação (elipse vazia) nas altas latitudes e nas áreas de reprodução (elipse preenchida) nas baixas latitudes, e as conexões migratórias entre essas áreas (linhas sólidas ou tracejadas). Áreas de alimentação (I à VI) e de reprodução (de A-G e X) divididas conforme à Comissão Internacional Baleeira (CIB).

As conexões migratórias entre as áreas de alimentação e reprodução das baleias jubarte eram presumidas durante o período da caça comercial (Kellogg 1929, Mackintosh

1946). No entanto, somente com a introdução das técnicas letais de marca-recaptura (*Discovery marks*) usadas durante à caça (Rayner 1940), é que os movimentos migratórios entre as áreas de reprodução e alimentação foram confirmados para algumas populações de baleias. No Hemisfério Sul, as primeiras evidências de conexão através desse método foram entre as jubartes da área de alimentação III e da costa de Madagáscar (Estoque reprodutivo C); e entre as jubartes das áreas IV e V, e das costas oeste e leste da Austrália (Estoques reprodutivos D e E, respectivamente) (Mackintosh 1942, Chittleborough 1965). Posteriormente, a técnica não-letal de marca-recaptura através da foto-identificação dos indivíduos (Katona & Whitehead 1981) confirmou a conexão migratória entre as jubartes da área de alimentação I e da área de reprodução nas costas do Equador, Colômbia, Panamá e Costa Rica (Estoque reprodutivo G) (Stone *et al*. 1990, Garrigue *et al*. 2002, Stevick *et al*. 2004, Rasmussen *et al*. 2007). Recentemente, a relação migratória entre as jubartes da área de reprodução na costa do Brasil (Estoque reprodutivo A) e da área de alimentação II foi demonstrada através da satélite telemetria e depois confirmada através dos dados de foto-identificação (Stevick *et al*. 2006, Zerbini *et al*. 2006, Engel & Martin 2009). Apesar desses métodos terem auxiliado na identificação das conexões migratórias entre as áreas de alimentação e reprodução para algumas populações de baleias jubarte, as relações migratórias para outras populações e o intercâmbio entre os estoques ainda permanecem desconhecidos.

Os avanços das análises genéticas nas últimas décadas e o uso de marcadores moleculares altamente variáveis, como o DNA mitocondrial (DNAmt) e os locos de microssatélites, tem auxiliado na definição e na identificação dos estoques para as baleias jubarte e para outras espécies de cetáceos (Hoelzel 1991). No Hemisfério Sul, recentes estudos genéticos sobre estrutura de população das áreas de reprodução das baleias jubarte nos oceanos Pacífico Sul, Atlântico Sul e Índico (Olavarria *et al*. 2007, Rosenbaum *et al*. 2009) foram realizados usando dados de DNA mitocondrial. Olavarria *et al*. (2007) encontraram uma diferenciação geográfica significativa entre os estoques reprodutivos do leste do oceano Índico (Estoque D – oeste da Austrália) e do oceano Pacífico Sul (Estoques E e F – Oceania e G – Colômbia). Devido à diferença genética entre as jubartes da Nova Caledônia e Tonga, o estoque E foi dividido em duas subpopulações (E2 e E3). Adicionalmente, o estoque F apresentou uma baixa, mas significante diferenciação entre as jubartes das ilhas Cook e Polinésia Francesa sugerindo a subdivisão do estoque em duas subpopulações (F1 e F2). A maior diferenciação encontrada dentro do Pacífico Sul foi entre a Côlombia (G) e a Oceania (E e F), sendo que o estoque G é o único representado pelo clado

AE, característico das populações do oceano Pacífico Norte. Esses resultados suportam o reconhecimento de 5 subpopulações de jubartes no oceano Pacífico Sul.

Do mesmo modo Rosenbaum *et al*. (2009) encontraram uma diferenciação populacional, baseada na variação de DNAmt, significativa entre os estoques dos oceanos Atlântico Sul (A – Brasil, e B – oeste da África) e Índico (C – leste da África, e X – norte do oceano Índico), bem como entre subpopulações entre esses estoques (entre B1 e B2; e entre C1 e C2, e C1 e C3). A maior diferenciação genética encontrada foi entre o estoque X e todos os outros, inclusive o estoque C localizado no mesmo oceano. Apesar da ausência de barreiras geográficas, a diferenciação através do DNAmt entre essas áreas de reprodução sugere que a dispersão das fêmeas é limitada. Reforçando a estrutura de população entres as áreas de reprodução, esses resultados sugerem uma transmissão maternal de longo tempo das rotas de migração e fidelidade às essas áreas de reprodução para as jubartes dos oceanos Pacífico Sul, Atlântico Sul, e Índico. No entanto, essa diferenciação é baixa e um grau de fluxo gênico ocorre entre esses estoques, sugerindo uma maior dispersão dos machos.

No Oceano Atlântico Sul Ocidental as baleias jubartes ocorrem ao longo da costa do Brasil (Figura 2), de aproximadamente 5° a 23° S (Zerbini *et al*. 2004, Andriolo *et al*. 2010), com avistagens ocasionais reportadas no Arquipélago de Fernando de Noronha (~ 3° S) e na costa sudeste e sul (até ~ 34° S) do país (Pinedo 1985, Lodi *et al*. 1994, Siciliano *et al*. 1999). Essa população de jubartes é reconhecida como o estoque reprodutivo A, e sua principal área de reprodução está localizada no Banco dos Abrolhos (16°40’ - 19°30’ S e 37°25’ - 39°45’ W), no sul da Bahia e norte do Espírito Santo (Siciliano *et al*. 1997, Martins *et al.* 2001, Freitas *et al.* 2004, Andriolo *et al.* 2006, 2010). Recentemente, um aumento das avistagens dos grupos de baleias jubarte tem sido reportado mais ao norte da Bahia (~ 500 km ao norte do Banco dos Abrolhos), principalmente no litoral de Salvador (Martins *et al*. 2001, Zerbini *et al*. 2004, Rossi-Santos *et al*. 2008). Além disso, os registros de encalhes ao norte e oeste de 5° S, nos estados do Ceará, do Piauí, do Maranhão e do Pará (Furtado-Neto *et al*. 1998, Siciliano *et al*. 2008, Magalhães *et al*. 2008, Meirelles *et al*. 2009, Pretto *et al*. 2009), e as avistagens nas proximidades das ilhas oceânicas do arquipélago de Trindade e Martim Vaz (Siciliano *et al*. 2012, Wedekin *et al*. *in press*) sugerem a recuperação da população e a reocupação da sua área de ocorrência histórica como indicado pelos dados da caça baleeira no Brasil (Paiva & Grangeiro 1965, Williamson 1975, Zerbini *et al*. 2004).

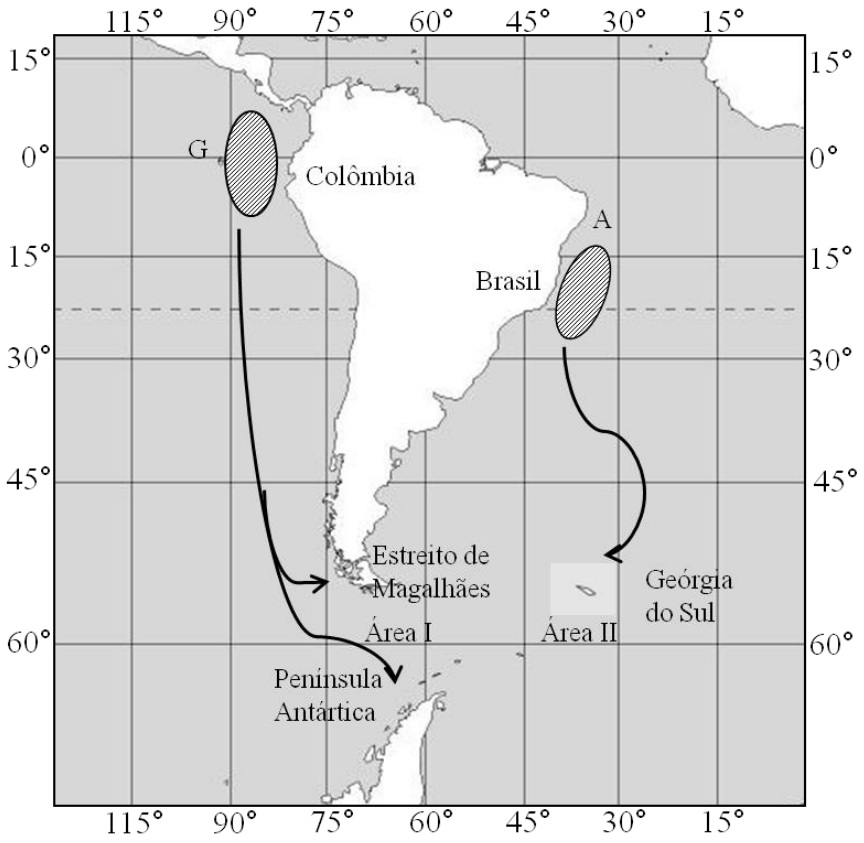



Figura 2: Estoques reprodutivos A e G, localizados respectivamente na costa leste e oeste da América do Sul, e suas respectivas áreas de alimentação, ilhas Geórgia do Sul e Sanduíche do Sul (Área II), Península Antártica (Área I) e Estreito de Magalhães.

As jubartes do estoque A migram para as áreas de alimentação em águas adjacentes às ilhas Geórgia do Sul e Sanduíche do Sul, no Mar da Escócia (Figura 2), entre as latitudes 54° e 60° S, e longitudes 33° e 22° W, sendo pertencente à Área II (Stevick *et al*. 2006b, Zerbini *et al*. 2006a, 2011, Engel *et al*. 2008, Engel & Martin 2009). Essa conexão migratória havia sido sugerida anteriormente por Slijper (1962, 1965) e Mackintosh (1965), mas nenhuma evidência direta tinha sido fornecida para suportar essa hipótese. Embora as técnicas de marca-recaptura (*Discovery marks*) usadas durante à caça comercial terem fornecido as primeiras evidências dos movimentos migratórios das baleias entre as áreas de reprodução e alimentação para algumas populações de jubartes, elas não foram bem sucedidas para as jubartes do Atlântico Sul Ocidental. Portanto, a primeira evidência da conexão entre o estoque A e as áreas de alimentação ao redor das ilhas Geórgia do Sul e Sanduíche do Sul foi através de duas jubartes monitoradas por satélite telemetria (Zerbini *et al*. 2006a), e posteriormente através de indivíduos foto-identificados no Banco dos Abrolhos e reavistados nas proximidades de Shag Rocks (oeste da Geórgia do Sul) e ilhas Sanduíche do Sul (Stevick *et*

*al*. 2006b, Engel & Martin 2009). Além disso, duas jubartes amostradas no ano de 2006 nas proximidades da ilha da Geórgia do Sul apresentaram haplótipos de DNA mitocondrial identificados na população brasileira (Engel *et al*. 2008), das quais uma apresentou uma possível relação mãe-filha com uma fêmea amostrada no Banco dos Abrolhos em 2001 (Cypriano-Souza *et al*. 2010).

No Oceano Pacífico Sul Oriental as baleias jubarte pertencentes ao estoque reprodutivo G ocorrem do norte da costa do Peru (6° S) a costa da Costa Rica (12° N) (Figura 2), com uma menor concentração de baleias ao redor do Arquipélago de Galápagos, localizado a 1000 km da costa do Equador (Acevedo-Gutiérrez & Smultea 1995, Flórez-González *et al*. 1998, Félix & Haase 2005, Félix *et al*. 2006, 2009, Pacheco *et al*. 2009). Esse estoque é considerado uma das populações de baleias jubarte do Hemisfério Sul, se reproduzindo durante o inverno austral, mas a maior parte da sua distribuição está localizada no Hemisfério Norte, alcançando a costa sul da América Central (Acevedo-Gutiérrez & Smultea 1995, Flórez-González *et al*. 1998, Rasmussen *et al*. 2007), aonde ocorre uma sobreposição espacial com as jubartes que se alimentam na costa da Califórnia e se reproduzem no inverno boreal nessa área (Calambokidis *et al*. 2000). Essa sobreposição espacial na mesma área de reprodução pode favorecer um possível fluxo gênico trans-equatorial entre as jubartes do Pacífico Norte e Sul, como tem sido sugerido pelo compartilhamento de algumas características genéticas entre essas populações (Baker *et al*. 1993, Acevedo-Gutiérrez & Smultea 1995, Medrano-González *et al*. 2001, Rasmussen *et al*. 2007).

A conexão migratória entre o estoque reprodutivo G e a área de alimentação em torno da Península Antártica (Figura 2), pertencente a Área I, tem sido confirmada através de reavistagens de indivíduos foto-identificados (Stone *et al*. 1990, Stevick *et al*. 2004, Rasmussen *et al*. 2007) e pela alta similaridade genética entre as jubartes dessas duas áreas (Olavarría *et al*. 2000, Caballero *et al*. 2001, Félix *et al*. 2012). Recentemente, o Estreito de Magalhães (Figura 2) foi também reconhecido como uma outra área de alimentação para as jubartes do estoque reprodutivo G (Gibbons *et al*. 2003, Acevedo *et al*. 2004, Acevedo *et al*. 2007, Capella *et al*. 2008). No entanto, dados de foto-identificação e genéticos têm demonstrado que as jubartes dessa área são distintas das baleias da área de alimentação na Península Antártica, sugerindo segregação de linhagens maternais do estoque reprodutivo G, similar aquela observada para as populações de baleias jubarte dos Oceanos Atlântico e Pacífico Norte (Acevedo *et al*. 2007, Olavarría *et al*. 2006).

Embora os estoques reprodutivos A e G, localizados respectivamente na costa leste e oeste da América do Sul, tenham apresentado um baixo compartilhamento de haplótipos de DNAmt (Olavarría *et al*. 2007, Engel *et al*. 2008), o primeiro registro de intercâmbio entre essas áreas foi evidenciado através de foto-identificação. Uma fêmea acompanhada pelo filhote avistada no Equador em 1996 foi posteriormente reavistada no Banco dos Abrolhos acompanhada por um indivíduo adulto em 1998 (Stevick *et al*. 2011). Apesar dessa primeira evidência de intercâmbio entre essas populações, e dos estudos anteriores terem fornecido importantes informações sobre ambos estoques reprodutivos, o grau de conectividade e de diferenciação entre essas populações precisa ser melhor investigado. Além disso, como os estudos anteriores usaram apenas dados da linhagem maternal (DNAmt), um estudo comparativo através de locos de microssatélites (biparentais) entre essas áreas é necessário para dar ou não suporte a distinção desses estoques reprodutivos. Portanto, a avaliação da estrutura populacional dos estoques A e G, e do nível de fluxo gênico entre essas populações, usando DNA mitocondrial e microssatélites como marcadores moleculares, é um dos objetivos dessa tese.

A caça comercial baleeira, principalmente durante o século XX, reduziu a população mundial de jubartes a menos de 10 % da original, antes do acordo de proteção internacional em 1966 (Tonnessen & Johnsen 1982). No Hemisfério Sul aproximadamente 200.000 baleias jubarte foram caçadas após o ano de 1900, principalmente pelas operações baleeiras em torno das áreas de alimentação na Antártica, reduzindo assim algumas populações a pequenas percentagens de seus tamanhos antes da exploração (Findlay 2001, Clapham & Baker 2002, Allison 2010). No Brasil as jubartes foram capturadas em pequena escala (~ 50 baleias por ano) principalmente na costa da Bahia entre os séculos XVII e XIX (Lodi 1992). Entretanto, as operações modernas, que tiveram início no século XX, expandiram as atividades das estações baleeiras principalmente para Costinha (7° S) entre 1910 e 1967, e para Cabo Frio (23° S) entre 1960 e 1963 (Paiva & Grangeiro 1965, 1970). No nordeste do Brasil uma média anual de 150 baleias foram capturada no período de 1911-1914/1924-1928, sendo que este número caiu para 12 indivíduos por ano de 1947 a 1963, refletindo a super exploração dessa população (Williamson 1975).

Após a proteção internacional, a frota da antiga União Soviética capturou ilegalmente baleias jubarte na costa central do Brasil até 1973 (Yablokov *et al.* 1998). O número total de baleias jubarte capturadas no Brasil é desconhecido, mas os dados do *Bureau of International Whaling Statistics* (BIWS) indicam que 1542 jubartes foram caçadas até 1963 (Williamson 1975). No entanto, os dados da caça no período entre 1929 e 1946 não foram considerados. O

tamanho da população do estoque reprodutivo A antes da caça comercial baleeira foi estimada usando dados dos registros de caça em aproximadamente 24.700 indivíduos, sendo que essa população foi reduzida a 2% de seu tamanho histórico até meados de 1950, quando havia menos do que 500 indivíduos (Zerbini *et al.* 2006b). Atualmente, a população vem se recuperando, com uma recente taxa de crescimento estimada em 7,4% por ano (Ward *et al*. 2011), e a mais recente estimativa de abundância em 2008 estimada em aproximadamente 9000 indivíduos (Wedekin *et al*. 2010). No entanto, a espécie se encontra listada no Apêndice I da CITES (*Convention on International Trade in Endangered Species of Wild Fauna and Flora*), na Lista Oficial de Espécies da Fauna Brasileira Ameaçadas de Extinção do IBAMA, e é considerada "menos preocupante" pela IUCN (*International Union for Conservation of Nature and Natural Resources*) e "vulnerável" pelo Plano de Ação de Mamíferos Aquáticos do Brasil (IBAMA 2001, IUCN 2008). Além disso, as principais ameaças atuais são o tráfego de grandes navios nas rotas desses animais, o emalhamento de filhotes em redes de pesca e as atividades petrolíferas na bacia de Campos (RJ) e adjacências causando preocupação quanto a futuros impactos sobre as baleias jubarte brasileiras (IBAMA, 2001).

O conhecimento do tamanho populacional histórico é importante para o manejo e a recuperação das espécies de cetáceos que foram muito caçadas. Estimativas de abundância das populações ou estoques de baleias pré-exploração e reconstrução da trajetória histórica de declínio e/ou recuperação são essenciais para avaliar o impacto da caça no ecossistema marinho e fornecer um embasamento na tomada de decisões presentes e futuras para recuperação dos estoques de baleias (Baker & Clapham 2004, Jackson *et al*. 2008). Dados genéticos têm fornecido informações para obter estimativas do tamanho histórico das populações de baleias (Rooney *et al.* 2001, Waldick *et al*. 2002, Roman & Palumbi 2003, Alter *et al.* 2007, 2012, Ruegg *et al.* 2010, 2013). No entanto, a alta diversidade genética encontrada em muitas populações de baleias tem gerado uma contradição entre as estimativas de abundância histórica dos registros da caça e dos dados genéticos. Além disso, estimativas genéticas do tamanho atual para a maioria dessas populações são desconhecidas, o que dificulta o entendimento da intensidade da redução causada pela exploração.

Dados genéticos de DNA mitocondrial das jubartes brasileiras apresentaram alta diversidade haplotípica e nucleotídica (Engel *et al.* 2008), de acordo com o descrito para outras áreas de reprodução estudadas nos oceanos do Hemisfério Sul (Olavarría *et al*. 2007, Rosenbaum *et al.*2009, Félix *et al*. 2012). Além disso, essa população apresentou alta diversidade nuclear e nenhum sinal de "*bottleneck*" genético foi detectado através dos dados de locos de microssatélites (Cypriano-Souza *et al.* 2010). Esses dados são consistentes com as

predições de que a caça comercial não reduziu suficientemente o tamanho das populações ou não durou gerações suficientes para significantemente reduzir a variabilidade genética das populações de baleias (Amos 1996). No entanto, esses estudos usaram métodos mais simples e com capacidade reduzida para detectar "*bottlenecks*" moderados. Além disso, nenhum estudo genético tem fornecido estimativas de tamanho efetivo e de censo para a população de jubartes do Oceano Atlântico Sul Ocidental. Deste modo, essa tese tem também como objetivo inferir a história demográfica dessa população através de estimativas do tamanho efetivo e de censo, atual e histórico, além de investigar as flutuações populacionais ao longo do tempo. Esses dados ajudarão a esclarecer as tendências atuais da população, determinando também o provável impacto do ecossistema sobre uma espécie ecologicamente importante e ameaçada.

# Capítulo 2 - Genetic diversity and effective and census population size of the humpback whales (*Megaptera novaeangliae*) wintering off Brazil (Breeding Stock A)

Ana Lúcia Cypriano-Souza, Tiago F. da Silva, Márcia H. Engel, Sandro L. Bonatto

**(Artigo a ser submetido à revista científica *Journal of Heredity*)**


"If we compare land animals in respect to magnitude, with those that take up their abode in the deep, we shall find they will appear contemptible in the comparison. The whale is doubtless the largest animal in creation."

— Oliver Goldsmith, *Goldsmith's Natural History*

**Genetic diversity and effective and census population size of the humpback whales (*Megaptera novaeangliae*) wintering off Brazil (Breeding Stock A)**

Ana Lúcia Cypriano-Souza[1,2]; Tiago Ferraz da Silva[1]; Márcia H. Engel[2] and Sandro L. Bonatto[1]

[1]*Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, PUCRS, Av. Ipiranga 6681, 90619-900 Porto Alegre, RS, Brazil*
[2]*Projeto Baleia Jubarte/Instituto Baleia Jubarte, Rua Barão do Rio Branco 26, 45900-000 Caravelas, BA, Brazil*

Address for correspondence: Dr. Sandro L. Bonatto, Faculdade de Biociências, PUCRS, Av. Ipiranga 6681, 90619-900 Porto Alegre, RS, Brazil, Phone: +55 (51) 3353.4717, Fax: +55 (51) 3320.3612, Email: slbonatto@pucrs.br

Running title: demographic history of the Brazilian humpback whales based on microsatellite loci

**Abstract**

Knowledge of recent and historical patterns of change in population size is essential for conservation. The anthropogenic changes in whale population sizes, in special before and during the peak of the commercial whaling are mostly studied by catch records. In the Southwestern Atlantic Ocean the humpback whale Breeding Stock "A" is distributed along the Brazilian coast (5° to 23° S) and it was reduced to nearly 2% of its historical size by commercial whaling during the 20$^{th}$ century. Here we genotyped 420 individuals of this population for ten microsatellite loci and estimate for the first time its contemporary and long-term effective population size ($N_e$) and the census size ($N_c$). We corroborated previous studies that found high genetic diversity for this population and an approximate Bayesian computation analysis highly supported a scenario of constant $N_e$ over size changes scenarios during the commercial whaling period. However, our estimates of contemporary $N_e$ at different time frames have detected a fluctuation of the population size during the whaling period (~ 2 to 4 generations ago). The $N_c$ between ~2,500 and 7,500 estimated here for the years around between 1980-1990 are broadly compatible with the most recent abundance surveys extrapolated to that time. Our estimate of long-term $N_c$ was larger than the pre-exploitation abundance suggesting that the Southwestern Atlantic humpback whale population may have been larger previously the start of modern whaling. Nevertheless, future studies should strive to reduce the uncertainties of several parameters for these estimates, including a much higher number of loci and markers such as SNPs.

*Keywords:* commercial whaling, bottleneck, effective size, microsatellites, demographic history

## Introduction

Drastic population size reductions or "bottlenecks" can lead to loss of genetic variation due to random genetic drift, inbreeding, and the resulting accumulation of deleterious mutations (Frankham et al. 2002, Lynch et al. 1995, Schwartz et al. 1998). Therefore, the detection of reductions in the census population size ($N_c$) and mainly in the effective population size ($N_e$, defined as the size of an ideal population experiencing the same rate of genetic change as the natural population under consideration (Crow and Kimura 1970)) is essential for conservation and management of threatened species (Luikart et al. 2010, Schwartz et al. 1998). Additionally, comparison of historic and contemporary $N_e$ is important for a better understanding the demographic history of the population, i.e., whether the population is declining or expanding (Leberg 2005, Ovenden et al. 2007). A fundamental contribution of population genetic theory to conservation biology has been the development of methods to estimate $N_e$ from allele frequency data (Leberg 2005, Luikart et al. 2010, Palstra and Ruzzante 2008, Schwartz et al. 1998, Tallmon et al. 2010, Waples 2005).

Knowledge of the historic population size is important for managing and restoring populations that suffered overexploitation, such as those of baleen whales. Estimation of the pre-exploitation abundance of whales populations or stocks and reconstruction of their demographic history (decline and/or recovery) are crucial to evaluate the impact of whaling on these species and the probability of current and future recovery of whale stocks (Baker and Clapham 2004, Jackson et al. 2008). Genetic data have provided important information to investigate historical fluctuations of whale population size (Alter et al. 2007, 2012, Roman and Palumbi 2003, Rooney et al. 1999, 2001, Ruegg et al. 2010, 2013, Waldick et al. 2002). However, as the majority of these studies have estimated the long-term effective population sizes, the contemporary effective sizes for these baleen whales are unknown, and thus hindering the comprehension of the exact magnitude of the level of reduction caused by exploitation.

Humpback whales (*Megaptera novaeangliae*, Borowski, 1781) were among the most exploited baleen whales by commercial whaling. They are found throughout the world's ocean basins undertaking annual migrations between the low latitude waters, where they breed and calve during the winter months, and the high latitude waters, where they feed during the summer (Dawbin 1966). Based on low latitude distributions in the Southern Hemisphere the International Whaling Commission (IWC) divided humpback whale stocks into seven breeding grounds (termed A-G) (IWC 2005). In the Southwestern Atlantic Ocean, humpback whales are presently found along the coast of Brazil (~ 5° to 23° S) (Andriolo et al.

2010). Distribution data have demonstrated that the main mating and calving ground of this population, recognized as the Breeding Stock A (BSA), is in the Abrolhos Bank (16°40'-19°30' S and 37°25'- 39°45' W) (Andriolo et al. 2006, 2010, Freitas et al. 2004, Martins et al. 2001, Siciliano et al. 1997), and an increasing number of sightings humpback whale groups, including females with calves, have been reported in the northern coast of Bahia, mainly off Salvador (Martins et al. 2001, Rossi-Santos et al. 2008, Zerbini et al. 2004). More recently, several evidences from satellite telemetry (Zerbini et al. 2006a, 2011), photo-identification (Engel and Martin 2009, Stevick et al. 2006), and genetic (Cypriano-Souza et al. 2010, Engel et al. 2008) data strongly support that the summer feeding grounds of this population are around South Georgia and South Sandwich islands in the Southern Ocean.

Commercial whaling during the 20th century reduced the global humpback whale population to an estimated less than 10% of the original size before worldwide protection in 1966 (Tønnessen and Johnsen 1982). In the Southern Hemisphere, approximately 200,000 humpback whales were hunted after 1900, mainly by whaling operations around Antarctica feeding areas, reducing some populations to small fractions of their pre-exploitation levels (Allison 2010, Clapham and Baker 2002, Findlay 2001). In Brazil, humpback whales were harvested off the southern and central coast in a small scale (~ 50 whales per year in Caravelas, Bahia) between the 17th and 19th centuries (Ellis 1969, Lodi 1992). However, modern whaling operations that began in the 20th century expanded the activities of the whaling stations mainly for the coasts of Costinha (7° S) (between 1910 and 1967) and Cabo Frio (23° S) (between 1960 and 1963), where 352 humpbacks were caught in 1913, and around 13 whales were caught in 1967, showing evidences of the reduction of this population in Brazil (Paiva and Grangeiro 1965, 1970, Williamson 1975). In addition, modern whaling activities in high-density areas in the Antarctic and Sub-Antarctic feeding grounds increased the annual catch to several thousand whales (Findlay 2001). Only in the surroundings of the South Georgia island about 19,000 humpback whales were killed between 1904 and 1913 (Headland 1984). Although the species have been protected since 1966, the former Soviet Union fleet took humpback whales illegally off the central coast of Brazil until 1973 (Yablokov et al. 1998).

The BSA population size before the exploitation by modern whaling was estimated using catch records to nearly 24,700 individuals, but it reached its lowest numbers in the late 1950s, when there were less than 500 individuals (Zerbini et al. 2006b). Presently this population is recovering fast (growth rate of 7.4% per annum, Ward et al. 2011) and the abundance in 2008 was estimated around 9,000 individuals (Wedekin et al. 2010).

Interestingly, despite these well documented census size changes in the BSA, recent studies have not detected a genetic bottleneck, that is, a significant reduction in the effective population sizes ($N_e$), of this population (Cypriano-Souza et al. 2010, Engel et al. 2008). This apparent incongruence is usually explained by both the short duration (in number of generations) and the relatively high absolute number of individuals during the bottleneck. However, these previous studies used simple methods with reduced power to detect moderate bottlenecks. Besides, none of the genetic studies so far have provided estimates of $N_e$ for the BSA population.

The present study aims to investigate the genetic diversity and the demographic history of the Brazilian humpback whale population based on the analysis of genotypes constructed from ten microsatellite loci for 420 individuals sampled at two geographic locations (Abrolhos Bank and Praia do Forte) off the Brazilian coast (Fig. 1). To accomplish this, we estimated the contemporary and long-term effective population size, and the census size, comparing the relationship between these estimates. We also evaluated the usefulness of different statistical estimators of contemporary effective population size.

## Materials and methods

### *Sample collection and DNA extraction*

Between 1999 and 2007, a total of 420 tissue samples of humpback whales were obtained along the coast of the Brazil. Most samples ($n$ = 379) were collected by the biopsy dart procedure (Lambertsen 1987) at two geographic locations off the Brazilian coast, the Abrolhos Bank ($n$ = 332), in the southern Bahia and northern Espírito Santo states, and Praia do Forte ($n$ = 47), northern Bahia. The remaining samples ($n$ = 41) resulted from individuals stranded along both states coast. Samples were preserved in 70% ethanol or DMSO, following Amos and Hoelzel (1990), and were stored at -20°C until processed. Genomic DNA was extracted using proteinase K digestion followed by phenol/chloroform extraction method (Palsbøll et al. 1995) or using a DNeasy tissue kit (QIAGEN).

### *Microsatellite genotyping*

Samples were screened for genetic variation at ten microsatellite loci (seven dinucleotides: EV1, EV37, EV94, EV96 (Valsecchi and Amos 1996), 199/200, 417/418, 464/465 (Schlötterer et al. 1991), and three tetranucleotides: GATA028, GATA053, GATA417, (Palsbøll et al. 1997)). Genotypes of 268 of the individuals used here were

described previously in Cypriano-Souza et al. (2010) and genotyping of the additional samples was conducted exactly as described in that study.

*Genetic variation*

The program MICRO-CHECKER v. 2.2.3 (Van Oosterhout et al. 2004) was used to identify possible null alleles, large allele dropout, and scoring errors due to stutter peaks. In the earlier study (Cypriano-Souza et al. 2010) locus 417/418 showed weak signs of null alleles, but here it was not discarded from the analyses because the homozygous excess was insufficient to suggest the presence of null alleles.

Genetic diversity was measured as the number of alleles per locus (*K*), observed and expected heterozygosities ($H_O$ and $H_E$, respectively) under Hardy-Weinberg equilibrium (HWE) (Nei 1978), using FSTAT v. 2.9.3 (Goudet 2002) and GENEPOP v. 4.1 (Raymond and Rousset 1995). FSTAT was also used to calculate Weir and Cockerham's (1984) $F_{IS}$. Deviations from HWE for each locus (Guo and Thompson 1992) and linkage disequilibrium between loci were tested using ARLEQUIN 3.5 (Excoffier and Lische 2010), adjusted for simultaneous comparisons with the sequential Bonferroni correction (Rice 1989).

*Demographic history*

Three methods were used to estimate contemporary effective population size ($N_e$). Two of these methods are based in the approximate Bayesian computation (ABC) approach, which uses summary statistics to estimate recent $N_e$. First, $N_e$ was estimated with ONeSAMP 1.2 (Tallmon et al. 2008, http://genomics.jun.alaska.edu) assuming ample upper and lower bounds (100-2000) on the prior distribution (uniform) for $N_e$. Eight summary statistics are calculated by ONeSAMP: number of alleles per locus (*A*), Garza-Williamson's index (*M*), expected heterozygosity ($H_E$), Wright's $F_{IS}$, the mean and variance of multilocus homozygosity, the difference of the natural logarithms of variance in allele length and heterozygosity, and the square of the correlation of alleles at different loci (Tallmon et al. 2008).

Additionally, the program DIYABC v. 1.0.46 (Cornuet et al. 2008, 2010) was used to test four different scenarios based in the possible demographic history of this humpback whale population during commercial whaling in the 20$^{th}$ century. Scenario 1 is a constant size population (no bottleneck), scenario 2 consisted of a population that is still experiencing a bottleneck, scenario 3 is a population that expanded recently from a bottlenecked population, and scenario 4 is a population that experienced a transitory bottleneck (Fig. 2a). The priors for

all parameters were uniformly distributed between specified minimum and maximum values (Table 1 and Fig. 2a) and were based on the available information of the whaling history of this population and its present day census data (see Introduction). The demographic parameters were: Scenario 1: $N_e$ (historical $N_e$); Scenario 2: $N_e2$ (current $N_e$), $N_a2$ (pre-bottleneck $N_e$); Scenario 3: $N_e3$ (current $N_e$), $N_a3$ ($N_e$ during bottleneck); $t$ (time since demographic change in scenarios 2 and 3); Scenario 4: $N_e4$ (current $N_e$), $N_b$ ($N_e$ during bottleneck), $N_a4$ (pre-bottleneck $N_e$), $t2$ and $t1$ (time since the beginning and end of the bottleneck, respectively). All times are in number of generations (generation time of 18 years taking into account the range between 12 and 24 years estimated for the humpback whales (Chittleborough 1965, Roman and Palumbi 2003)) from the present, with $t2 > t1$. The 10 microsatellites loci were assumed to evolve under the generalized stepwise mutation model (GSM) (Estoup et al. 2002) with a widely used mutation rate ($\mu$) range for mammals from $10^{-4}$ to $10^{-3}$ per generation (Ellegren 1995, Hoffman et al. 2011, Whittaker et al. 2003) and the coefficients of geometric distribution ($P$) from 0.1 to 0.7. Motif sizes and alleles ranges were adjusted to 10 loci. The summary statistics were the mean number of alleles ($A$), genetic diversity ($H_E$), allelic size range ($AR$), and Garza-Williamson's index ($M$). A total of 3,000,000 simulations were performed to generate the reference table, using the four scenarios according to their prior probability and their parameter values drawn from the prior distributions. The posterior probability of each scenario was assessed using both direct estimate and logistic regression approaches using the 500 and 30,000 best simulations respectively. Under an ABC approach, the best scenario is the one with the simulated data set closest to observed data set. For each scenario, the posterior distribution of the parameters were estimated using logit transformation for the 8,000 best simulations.

Alternatively, the program $LDN_E$ (Waples and Chi Do 2008) was used to estimate $N_e$ from genotypic data based on the linkage disequilibrium (LD) method with the bias correction developed by Waples (2006). The program calculates separate estimates using different criteria for excluding rare alleles. We used the random mating model, jackknife methods for obtaining confidence intervals to $N_e$ and the following critical values ($P_{crit}$): 0.05; 0.02; 0.01. Finally, estimates of long-term effective population size ($N_e$) were made using a moment estimator, which assumes mutation drift equilibrium (Leberg 2005). We used expected heterozygosity ($H_E$) levels at individual microsatellite loci under the stepwise (SMM) mutation model (Mitrovski et al. 2007, Rooney et al. 1999, Waldick et al. 2002) as:

SMM: $N_e = \{(1/1\text{-}H)^2 - 1\}/8\mu$

where $N_e$ is the effective population size, $H$ is the mean expected heterozygosity and $\mu$ is the mutation rate per locus. We used the limits used for the DIYABC method ($10^{-4}$ to $10^{-3}$ per generation) as the mutation rate.

We calculated the total census size ($N_c$) from the effective size ($N_e$) by multiplying by a conservative 2:1 ratio of total adults to breeding adults ($N_T$:$N_e$) (Alter et al. 2007, 2012, Roman and Palumbi 2003, Ruegg et al. 2010, 2013), and by the proportion of juveniles in the population (number of adults + juveniles)/(number of adults), estimated between 1.6 to 2.0 for humpback whales (Chittleborough 1965, Roman and Palumbi 2003). Therefore, the average ratio of census population size to effective population size estimated was 3.6, with a variation from 3.2 to 4.0.

## Results

### *Genetic variability*

Individual multilocus genotypes were on average 98.5 % complete. Summary statistics are presented in Table 2. The number of alleles identified at the ten microsatellite loci ranged from 5 (EV1) to 18 (GATA417) with a mean of 12.6. The mean observed ($H_O$) heterozygosity was 0.736, ranging from 0.553 (EV1) to 0.923 (GATA417), and the mean expected ($H_E$) heterozygosity was 0.746, ranging from 0.532 (EV1) to 0.923 (EV37). Population-wide $F_{IS}$ values were low for all loci (below 0.05), except for the locus GATA053 ($F_{IS}$ = 0.053) and the locus 417/418 ($F_{IS}$ = 0.068), but these values were not significant. Moreover, no evidence of null alleles and no significant deviation from HWE expectations were seen at any of the loci. Pairwise comparison of allele frequencies revealed no significant linkage disequilibrium after Bonferroni correction.

### *Demographic history*

Effective ($N_e$) and total census ($N_c$) population sizes estimates based on the different methods are shown in Table 3. In the comparison of the four scenarios (constant population, bottlenecked, expanded, and transitory bottleneck) using the ABC approach implemented in DIYABC, the constant population (no demographic changes) scenario for both, direct estimate and logistic regression approaches (Fig. 2b and 2c) is highly supported (posterior probability > 0.99) in relation to the other scenarios with a bottleneck. In the constant scenario, the mode of the posterior distribution for $N_e$ was 4,170 (95% CI = 2,330 - 26,600)

(Fig. 3). The mode for the $N_c$ was then 15,012 (95% CI = 8,388 - 95,760), using the 3.6 census/effective ratio.

The other methods used to determine the contemporary $N_e$ provided different estimates. While those reported by ONeSAMP were lower, with a mean of 560 individuals (95% CI = 475 - 731), those obtained with $LDN_E$ ranged from 1,029 ($P_{crit}$ = 0.01, 95% CI = 698 - 1,829) to 1,524 individuals ($P_{crit}$ = 0.05, 95% CI = 682 - ∞) for the different critical values. However, $P_{crit}$ = 0.02 is indicated to provide better precision, therefore our more reliable estimate was 1,061 whales (95% CI = 688 - 2,104). The results obtained with ONeSAMP were robust to changes in the bounds for the $N_e$ prior,where the mean $N_e$ of 540 (95% CI = 490 - 617) and 510 (95% CI = 458 - 585) were obtained for the priors limits of 200-1000 and 100-800, respectively. The total census population size ($N_c$) was then 2,016 for $N_e$ estimated with ONeSAMP (100-2000 prior) and 3,820 for $N_e$ estimated with $LDN_E$ ($P_{crit}$ = 0.02).

The long-term $N_e$ and census size estimates based on the $H_E$ with SMM mutation model calculated based on different mutation rates were shown in Table 3. Using an average mutation rate of 2 X $10^{-4}$ (Hoelzel et al. 2007), $N_e$ and $N_c$ were estimated at 28,238 and 101,656, respectively.

**Discussion**

Our extended sampling corroborates our previous results (Cypriano-Souza et al. 2010) on the high nuclear diversity of the humpback whale population that winters off the Brazilian coast (BSA) and is compatible with its high mtDNA variability (Engel et al. 2008). This high genetic diversity is in agreement with other breeding grounds studied in the Southern Hemisphere for both nuclear and mitochondrial markers (e.g. Garrigue et al. 2004, Olavarría et al. 2007, Pomilla and Rosenbaum 2006, Rosenbaum et al. 2009, Valsecchi et al. 2002). However, in the Southwestern Atlantic population, as well as most other humpback populations, severe reductions of their historical size by commercial whaling are well documented. The lowest number reached for stock A was in the late 1950s, when around 500 individuals (95% CI = 152 to 3,687) were estimated to have existed in the population (Zerbini et al. 2006b). On the other hand, our previous study did not detect any significant signal of a genetic bottleneck in this population (Cypriano-Souza et al. 2010) using three standard methods: heterozygosity excess (Cornuet and Luikart 1996), mode-shift (Luikart et al. 1998) and M-ratio tests (Garza and Williamson 2001). This result is corroborated here with the extended data set and an approximate Bayesian computation approach using coalescence

simulations, in which by far the best supported scenario was a constant population compared with those in which a population experienced a single size change (expansion or bottleneck) or a bottleneck during the commercial whaling (between 2 and 8 generations ago) followed by an expansion (Fig. 2).

As discussed previously for the Brazilian population (Cypriano-Souza et al. 2010, Engel et al. 2008), these results are consistent with the hypothesis (Amos 1996) that the commercial whaling did not last for enough generations or did not sufficiently reduce the population size to significantly reduce its genetic diversity. The magnitude of the genetic bottleneck is dependent on both the absolute size of the population during the bottleneck and its duration (Frankham et al. 2002). Therefore shorter bottlenecks with a not so small effective population size left weaker signals and are therefore more difficult to detect. More recently, Phillips et al. (2012) showed a similar result with an ABC analysis for the bowhead whales (*Balaena mysticetus*) which did not support a bottleneck scenario. In contrast, recently Hoffman et al. (2011) have not detected bottleneck for the Antarctic fur seal (*Arcthocephalus gazella*) using the standard tests, but their ABC analysis supported a bottleneck scenario. For these different results based on ABC approach, it was suggested that recent population size changes is not sufficient to drive genetic signal in a species with long generation time (~ 50 years for bowhead whales) (Phillips et al. 2012). However, Alter et al. (2012) detected a recent bottleneck (at 93 years ago, ~ 6 generations) using an ABC approach for the eastern Pacific gray whales (*Eschrichtius robustus*), but this whale species has a shorter generation time (~ 15 years) and they included ancient samples in the analysis. This was the first study that detected a bottleneck for a target species of the whaling.

Although the ABC test did not show evidence of a significant genetic bottleneck during the historical whaling (Fig. 2), our $N_e$ estimates based on different methods may give us clues that this process may have left some signal in the genetic pool of the population that may not be detected by any single method. The three different methods of contemporary $N_e$ estimation used here, besides having some different assumptions, are also known to estimate $N_e$ on different time frames (Table 3). $LDN_E$ estimates $N_e$ of the previous generation (Waples 2006), ONeSAMP from the previous two to eight generations (Tallmon et al. 2008); while $N_e$ estimates based on DIYABC depend of the scenario supported for which $N_e$ is being estimated (see details on the original descriptions for each method listed in Materials and Methods). In this context, the smallest contemporary $N_e$ estimated by ONeSAMP in our data may reflects the period of overexploitation of this stock between 1904 to 1967, when commercial whaling was most extensive (Paiva and Grangeiro 1970). This corresponds to

approximately two to four generations ago (relative to the generation time estimated for this species of 12-24 years; Chitteborough 1965, Roman and Palumbi 2003). Furthermore, a higher $N_e$ estimated by LDNe than the ONeSAMP estimate is expected since the former represent a more recent generation and the increase in the population in this area since the end of the whaling is well documented (see below). Finally, considering the scenario of constant population, the larger $N_e$ (~ 4,200) estimated by DIYABC seems to reflect the larger long-term pre-whaling size of this population (see below, Zerbini et al. 2006b). In addition, our long-term $N_e$ based on moment estimator using SMM also supports a much larger historical abundance of this population. However, for this latter the accuracy of the estimate depends on the mutational model assumed for microsatellite loci under study (stepwise), and the assumed mutation rate at these loci (Leberg 2005, Rooney et al. 1999). Consequently, more loci, more realistic scenarios (non-instantaneous population growth, gene flow, etc.) and methods are necessary to better estimate the demographic history of this population.

There have been recently several contemporary abundance estimates of the Brazilian humpback whale population based on different methods. An estimate based on photo-identification data collected between 1996 and 2000 in the Abrolhos Bank suggested 2,393 whales (95% CI = 1,924-3,060) using the multiple-recapture model of closed population, 3,871 (95% CI = 2,745-5,542) by the open population model, and 3,000 (95% CI = 2,500-3,650) by the Whitehead's model allowing for emigration and re-immigration (Freitas et al. 2004). Andriolo et al. (2006) employed aerial surveys between the north limit of Bahia (12°10' S) and the southern limit of Espírito Santo (20°42' S) in 2001 and estimated a population size of 2,229 whales (95% CI = 1,201-4,137), with corrected analysis for the surface time. However, the above estimates were based on data collected in an area corresponding to only a portion of the known wintering grounds. Only the most recent estimates of 6,404 individuals (95% CI = 5,085–8,068) in 2005 and of 9,330 individuals (95% CI = 7,185–13,214) in 2008, were derived from aerial surveys (Andriolo et al. 2010, Wedekin et al. 2010) that covered the entire stock range. The census size ($N_c$) estimated by the $LDN_E$ method was between around 2,500 and 7,500 (Table 3) corresponding to the previous generation of our sample (obtained between 1999 and 2007), which given the uncertainty in the generation time would roughly correspond to the years between around 1980 and 1990. These values are broadly compatible with the most recent abundance survey by Wedekin et al. (2010) extrapolated to the past using a growth rate of 7.4% per annum (Ward et al. 2011). Similarly, the ONESAMP estimate between 1,700 and 2,924 correspond to a $N_c$ between two to eight generations before the sampling, which is consistent with the estimated lowest

abundance number reached in the 1950s (~ 3 generations before sampling) between 152 and 3,687 individuals.

The most recent and complete study on historical abundance of the stock A based on catch records using a Bayesian statistical method estimated the population size nearly 24,700 individuals (95% CI = 22,804-31,220) before exploitation by modern whaling (Zerbini et al. 2006b). Our point estimate for recent pre-whaling $N_c$ (~ 15,000) using the ABC approach was smaller than the pre-whaling abundance cited above, although the confidence intervals widely overlap. In contrast, the long-term $N_c$ estimate (using a moment estimator (SMM)) of about 101,000 individuals was much larger (~ 4 times) than the BSA pre-whaling abundance (24,700 individuals) based on catch records. However, this estimate was similar to the two more recent estimates for the humpback whales from North Atlantic. Alter and Palumbi (2009) updated the original mtDNA-based estimate (~ 240,000 individuals, Roman and Palumbi 2003) using a more accurate mutation estimate for which they estimated 150,000 individuals. More recently, Ruegg et al. (2013) estimated a long-term population size of 112,000 individuals in the North Atlantic, based on a multi-locus estimate.

Since this is the first estimate from the Southwestern Atlantic population size using genetic data there is no previous measure to compare. Although a quantitative comparison between the estimates from other populations and species and our estimates is difficult since the data type (mtDNA or nuclear introns vs. microsatellites) and methods were different, all these estimates of historical abundances have suggested that the whale populations were much larger than previously estimated by catch records. However, a direct connection between pre-whaling abundance data with long-term genetic census estimates may be unwarranted. First, most results present broad intervals usually derived from large uncertainties of several parameters, such as generation time, mutation rate, or the relation between $N_c$ and $N_e$. Most importantly, these estimates represent different time frames, since pre-exploitation abundance represent population size at the start of the whaling records (in the case of BSA stock, less than a century) while the long-term genetic estimates represent the weighted harmonic mean of population sizes over $4N_e$ generations, and therefore it is influenced by the demographic dynamics of the population (see Charlesworth 2009 for a review).

Overall, our results corroborate the high genetic diversity of the BSA and previous studies that do not found statistically significant reduction in this genetic diversity caused by modern whaling (Cypriano-Souza et al. 2010, Engel et al. 2008). However, the different methods for contemporary $N_e$ estimates seems have detected a fluctuation of the population size during the commercial whaling (~ 2 to 4 generations ago). These results suggest that

using methods that estimated effective population size at different time frames may be an efficient approach to detect and test recent anthropogenic demographic changes. Additionally, the smaller genetic-based estimate of the recent long-term census size than the pre-exploitation abundance suggests that this estimate corresponds to a $N_c$ after the onset and before the peak of whaling, indicating the begin of a decline in this population. Finally, the estimate of the long-term historical census size also suggests that the Southwestern Atlantic humpback whale population was larger than the pre-exploitation abundance corroborating the long-term $N_c$ in other stocks. These latter two results however still present large confidence intervals derived from uncertainties of several key parameters from the genetic estimates, and future studies should therefore strive to reduce these uncertainties, for example with a much higher number of loci and markers such as SNPs.

**Funding**

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [to A.L.C.-S.]; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [to S.L.B.]; Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) [to S.L.B.]; Instituto Chico Mendes de Conservação da Biodiversidade/Ministério do Meio Ambiente (ICMBIO/MMA); and Petróleo Brasileiro S.A (PETROBRAS) [to Instituto Baleia Jubarte]. Biopsy samples were collected off Brazil under permit SISBIO 21489-1.

**Acknowledgements**

We thank everyone from the staff of Instituto Baleia Jubarte involved in biopsy sampling and in the collection of samples from whales stranded, specially Milton Marcondes for organization and shipment of the samples. We also thank the colleagues of the Laboratório de Biologia Genômica e Molecular/PUCRS for their help in lab activities, specially Cladinara Sarturi for laboratory assistance. We are grateful to Taran Grant for his suggestions on an earlier version of this paper.

**References**


Alter SE, Rynes E, Palumbi SR. 2007. DNA evidence for historic population size and past ecosystem impacts of gray whales. Proc Natl Acad Sci USA. 104:15162-15167.

Alter SE, Palumbi SR. 2009. Comparing evolutionary patterns and variability in the mitochondrial control region and cytochrome b in three species of baleen whales. J Mol Evol. 68:97-111.

Alter SE, Newsome SD, Palumbi SR. 2012. Prewhaling genetic diversity and population ecology in eastern pacific grey whales: insights from ancient DNA and stable isotopes. PLoS ONE. 7:e35039.

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

Amos B. 1996. Levels of genetic variability in cetacean populations have probably changed little as a result of human activities. Rep Int Whaling Comm. 46:657-658.

Andriolo A, Martins CCA, Engel MH, Pizzorno JL, Más-Rosa S, Freitas A, Morete ME, Kinas PG. 2006. The first aerial survey to estimate abundance of humpback whale (*Megaptera novaeangliae*) in the breeding ground off Brazil. J Cetacean Res Manage. 8:307-311.

Andriolo A, Kinas PG, Engel MH, Martins CCA. 2010. Humpback whale population estimates and distribution along the Brazilian breeding ground. Endang Spec Resear. 11:233-243.

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

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics. 10:195-205.

Chittleborough RG. 1965. Dynamics of two populations of the humpback whale, *Megaptera novaeangllae* (Borowski). Aust J Mar Freshw Res. 16:33-128.

Clapham PJ, Baker CS. 2002. Modern whaling. In: Perrin WF, Würsig B, Thewissen JGM, editors. Encyclopedia of Marine Mammals, New York: Academic Press; p. 1328-1332.

Cornuet JM, Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. 144:2001–2014.

Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, Estoup A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics. 24(23):2713-2719.

Cornuet JM, Ravigné V, Estoup A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1. 0). Bmc Bioinformatics. 11(1):401.

Crow JF, Kimura M. 1970. An introduction to population genetics theory. Minneapolis: Burgess.

Cypriano-Souza AL, Fernández GP, Lima-Rosa CAV, Engel MH, Bonatto SL. 2010. Microsatellite genetic characterization of the humpback whale (*Megaptera novaeangliae*) Breeding Ground off Brazil (Breeding Stock A). J Hered. 101(2):189-200.

Dawbin W. 1966. The seasonal migratory cycle of the humpback whale. In: Norris KS, editor. Whales, Dolphins and Porpoises. Berkeley: University of California Press; p. 145-170.

Ellegren H. 1995. Mutation rates at porcine microsatellite loci. Mamm Genome. 6:376–377.

Ellis M. 1969. A baleia no Brasil colonial. Edições Melhoramentos, Editora da Universidade de São Paulo, São Paulo. 235 pp.

Engel MH, Fagundes NJR, Rosenbaum HC, Leslie MS, Ott PH, Schmitt R, Secchi E, Dalla Rosa L, Bonatto SL. 2008. Mitochondrial DNA diversity of the Southwestern Atlantic humpback whale (*Megaptera novaeangliae*) breeding area off Brazil, and the potential connections to Antartic feeding areas. Conserv Genet. 9:1253-1262.

Engel MH, Martin AR. 2009. Feeding grounds of the western South Atlantic humpback whale population. Mar Mamm Sci. 25:964–969.

Estoup A, Jarne P, Cornuet JM. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol. 11:1591-1604.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Lynux and Windows. Mol Ecol Resources, 10:564-567.

Findlay KP. 2001. A review of humpback whale catches by modern whaling operations in the Southern Hemisphere. Mem Quensl Mus. 47(2):411-20.

Frankham R, Ballou JD, Briscoe DA. 2002. Introduction to Conservation Genetics. Cambridge University Press.

Freitas AC, Kinas PG, Martins CCA , Engel MH. 2004. Abundance of humpback whales on the Abrolhos Bank wintering ground, Brazil. J Cetacean Res Manage. 3:225-230.

Garrigue C, Dodemont R, Steel D, Baker CS. 2004. Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm low abundance and reproductive

autonomy of numpback whales on the wintering grounds of New Caledonia. Mar Ecol Progr Ser. 274:251-262.

Garza JC, Williamson EG. 2001. Detection of reduced population size using data from microsatellite loci. Mol Ecol. 10:305-318.

Goudet J. 2002. FSTAT: a program to estimate and test gene diversities and fixation indices. Lausanne: Institute of Ecology.

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

Headland R. 1984. The island of South Georgia. London: Cambridge University Press; p 291.

Hoelzel AR, Hey J, Dahlheim ME, Nicholson C, Burkanov V, Blacks N. 2007. Evolution of population structure in a highly social top predator, the killer whale. Mol Biol Evol. 24:1407-1415.

Hoffman JI, Grant SM, Forcada J, Phillips CD. 2011. Bayesian inference of a historical bottleneck in a heavily exploited marine mammal. Mol Ecol. 20(19):3989-4008.

International Whaling Commission, 2005. Report of the Scientific Committee. Annex H. Report of the sub-committee on other Southern Hemisphere whale stocks. J Cetacean Res Manage. 7 (Suppl.):235-46.

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.

Lambertsen RH. 1987. A biopsy system for large whales and its use for cytogenetics. J Mammal. 68:443-445.

Leberg P. 2005. Genetic approaches for estimating the effective size of populations. J Wild Manage. 69:1385-1399.

Lodi L. 1992. Uma história da caça à baleia. Ciência Hoje 14(81):78-83.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered. 89:238-247.

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW. 2010. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches**.** Conserv Genet. 11:355-373.

Lynch M, Conery J, Burger R. 1995. Mutation, accumulation and the extinction of small populations. Amer Natur. 146:489-518.

Martins CCA, Morete ME, Engel MH, Freitas A, Secchi ER, Kinas PG. 2001. Aspects of habitat use patterns of humpback whales in the Abrolhos bank, Brazil, breeding ground. Mem Queensl Mus. 47:563-570.

Mitrovski P, Heinze DA, Broome L, Hoffmann AA, Weeks AR. 2007. High levels of variation despite genetic fragmentation in populations of the endangered mountain pygmy-possum, *Burramys parvus*, in alpine Australia. Mol Ecol. 16:75-87.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89:583-590.

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 humpback whales throughout the South Pacific and the origin of the eastern Polynesian breeding grounds. Mar Ecol Prog Ser. 330:257-268.

Ovenden JR, Peel D, Street R, Courtney AJ, Hoyle SD, Peel S, Podlich H. 2007. The genetic effective and adult census size of an Australian population of tiger praws (*Penaeus esculentus*). Mol Ecol. 16:127-138.

Paiva MP, Grangeiro BF. 1965. Biological investigations on the whaling seasons 1960-1963, off the northeastern coast of Brazil. Arq Estac Biol Mar Univ Ceara. 5:29-64.

Paiva MP, Grangeiro BF. 1970. Investigations on the whaling seasons 1964-1967, off the northeastern coast of Brazil. Arquivos de Ciências do Mar. 10:111-126.

Palsbøll PJ, Clapham PJ, Mattila DK, Larsen F, Sears R, Siegismund HR, Sigurjónsson J, Vásquez O, Artander P. 1995. Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. Mar Ecol Prog Ser. 116:1-10.

Palsbøll PJ, Bérubé M, Jorgensen H. 1997. Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. Mol Ecol. 6:893-895.

Palstra FP, Ruzzante DE. 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? Mol Ecol. 17:3428-3447.

Phillips CD, Hoffman JI, George JC, Suydam RS, Huebinger RM, Bickham JW, Patton JC. 2012. Molecular insights into the historic demography of bowhead whales: understanding the evolutionary basis of contemporary management practices. Ecol Evol. 3(1):18-37.

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

Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered. 86:248-249.

Rice WR. 1989. Analyzing table of statistical tests. Evolution. 43:223-225.

Roman J, Palumbi S. 2003. Whales before whaling in North Atlantic. Science. 301:508-510.

Rooney AP, Honeycutt RL, Davis SK, Derr JN. 1999. Evaluating a putative bottleneck in a population of bowhead whales from patterns of microsatellite diversity and genetic disequilibria. J Mol Evol. 49:682-690.

Rooney AP, Honeycutt RL, Derr JN. 2001. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. Evolution. 55:1678-1685.

Rosenbaum HC, Pomilla C, Mendez M, Leslie MS, Best PB, Findlay KP, Minton G, Ersts PJ, Collins T, Engel MH, Bonatto SL, Kotze DPGH, Meÿer M, Barendse J, Thornton M, Razafindrakoto Y, Ngouessono S, Vely M, Kiszka J. 2009. Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. PLoS ONE. 4:e7318.

Rossi-Santos MR, Neto ES, Baracho CG, Cipolotti SR, Marcovaldi E, Engel MH. 2008. Occurrence and distribution of humpback whales (Megaptera novaeangliae) on the North coast of the State of Bahia, Brazil, 2000-2006. ICES J Mar Sci. 65:1-7.

Ruegg KC, Anderson EC, Baker CS, Vant M, Jackson JA, Palumbi SR. 2010. Are Antarctic minke whales unusually abundant because of 20th century whaling? Mol Ecol. 19:281-291.

Ruegg KC, Rosenbaum HC, Anderson EC, Engel MH, Rothchild A, Baker CS, Palumbi SR. 2013. Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure. Conserv Genet. 14:103-114.

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

Schwartz MK, Tallmon DA, Luikart G. 1998. Review of DNA-based census and effective population size estimators. An Conserv. 1:293-299.

Siciliano S. 1997. Características da população de baleias-jubarte (*Megaptera novaeangliae*) da costa brasileira, com especial referência aos Bancos de Abrolhos. MS Thesis, Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro.

Tallmon DA, Koyuk A, Luikart G, Beaumont MA. 2008. ONESAMP: a program to estimate effective population size using approximate Bayesian computation. Mol Ecol Resources. 8:299-301.

Tallmon DA, Gregovich D, Waples RS, Baker CS, Jackson J, Taylor BL, Archer E, Martien KK, Allendorf FW, Schwartz MK. 2010. "When are genetic methods useful for estimating contemporary abundance and detecting population trends?." Mol Ecol Resources 10: 684-692.

Tønnessen JN, Johnsen AO. 1982. The History of Modern Whaling. Berkeley: University of California Press.

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

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

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping erros in microsatelite data. Mol Ecol Notes. 4:535-538.

Waldick RC, Kraus S, Brown M, White BN. 2002. Evaluating the effects of historic bottleneck events: an assessment of microsatellite variability in the endagered, North Atlantic right whale. Mol Ecol. 11:2241-2249.

Waples RS. 2005. Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? Mol Ecol. 14:3335-3352.

Waples RS. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conserv Genet. 7:167-184.

Waples RS, Chi Do, 2008. $LDN_E$: A program for estimating effective population size from data on linkage disequilibrium. Mol Ecol Resources. 8:753-756.

Ward E, Zerbini, AN, Kinas PG, Engel MH, Andriolo A. 2011. Estimates of population growth rates of humpback whales (*Megaptera novaeangliae*) in the wintering grounds off the coast of Brazil (Breeding Stock A). J Cet Res Manage. 3:145-149.

Wedekin LL, Engel MH, Azevedo A, Kinas PG, Andriolo A, Simões-Lopes PC. 2010. Density and abundance of the humpback whale in the Brazilian breeding ground (stock A): aerial survey, 2008. Unpublished paper SC/62/SH28 presented to the Scientific Committee of the International Whaling Commission.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the Analysis of Population Structure. Evolution. 38:1358-1370.

Williamson GR. 1975. Minke whales off Brazil. Scientific Reports of the Whales Research Institute. 27:37-59.

Whittaker JC, Harbord RM, Boxall N, Mackay I, Dawson G, Sibly RM. 2003. Likelihood-based estimation of microsatellite mutation rates. Genetics. 164:781-787.

Yablokov AV, Zemsky VA, Mikhalev YA, Tornosov VV, Berzin AA. 1998. Data on Soviet whaling in the Antarctic in 1947-1972 (population aspects). Russ J Ecol. 29:38-42.

Zerbini AN, Andriolo A, da Rocha JM, Simões-Lopes PC, Siciliano S, Pizzorno JL, Waite JM, DeMaster DP, VanBlaricon GR. 2004. Winter distribution and abundance of humpback whales (*Megaptera novaeangliae*) off Northeastern Brazil. J Cetacean Res Manage. 6:101-107.

Zerbini AN, Andriolo A, Heide-Jorgensen MP, Pizzorno JL, Maia YG, VanBlaricom GR, DeMaster DP, Simões-Lopes PC, Moreira S, Bethlem C. 2006a. Satellite-monitored movements of humpback whales *Megaptera novaeangliae* in the Southwest Atlantic Ocean. Mar Ecol Prog Ser. 313:295-304.

Zerbini AN, Ward E, Kinas PG, Engel MH, Andriolo A. 2006b. A Bayesian Assessment of the Conservation Status of Humpback Whales (*Megaptera novaeangliae*) in the Western South Atlantic Ocean. Unpublished paper SC/58/SH2 presented to the Scientific Committee of the International Whaling Commission.

Zerbini AN, Andriolo A, Danilewicz D, Heide-Jorgensen MP, Gales N, Clapham PJ. 2011. An update on research on migratory routes and feeding destinations of Southwest Atlantic humpback whales. Unpublished paper SC/63/SH23 presented to the Scientific Committee of the International Whaling Commission.

**Figure Legends**

Figure 1. Map of the surveyed areas, showing the geographic locations of the two sampling sites (zoom) of the humpback whale breeding ground off the Brazilian coast (BSA).

Figure 2. a) The four demographic scenarios tested with the DIYABC approach: 1 - constant population, 2 - bottlenecked population, 3 - expanded population, 4 - population with a transitory bottleneck. Demographic parameters: $N_e$ - historical; $N_e2$, $N_e3$ and $N_e4$ - current; $N_a2$ and $N_a4$ - pre-bottleneck; $N_a3$ and $N_b$ - during bottleneck. b) Posterior probability (y-axis) of the four scenarios in different numbers of selected closest-to-observed simulations based on the direct estimate. The posterior probability of each scenario is given at the bottom. c) Logistic regression.

Figure 3. Posterior distribution (in green) of the parameter $N_e$ from the best supported scenario (Scenario 1, constant population) as estimated in the program DIYABC. The red line is the prior distribution for $N_e$.

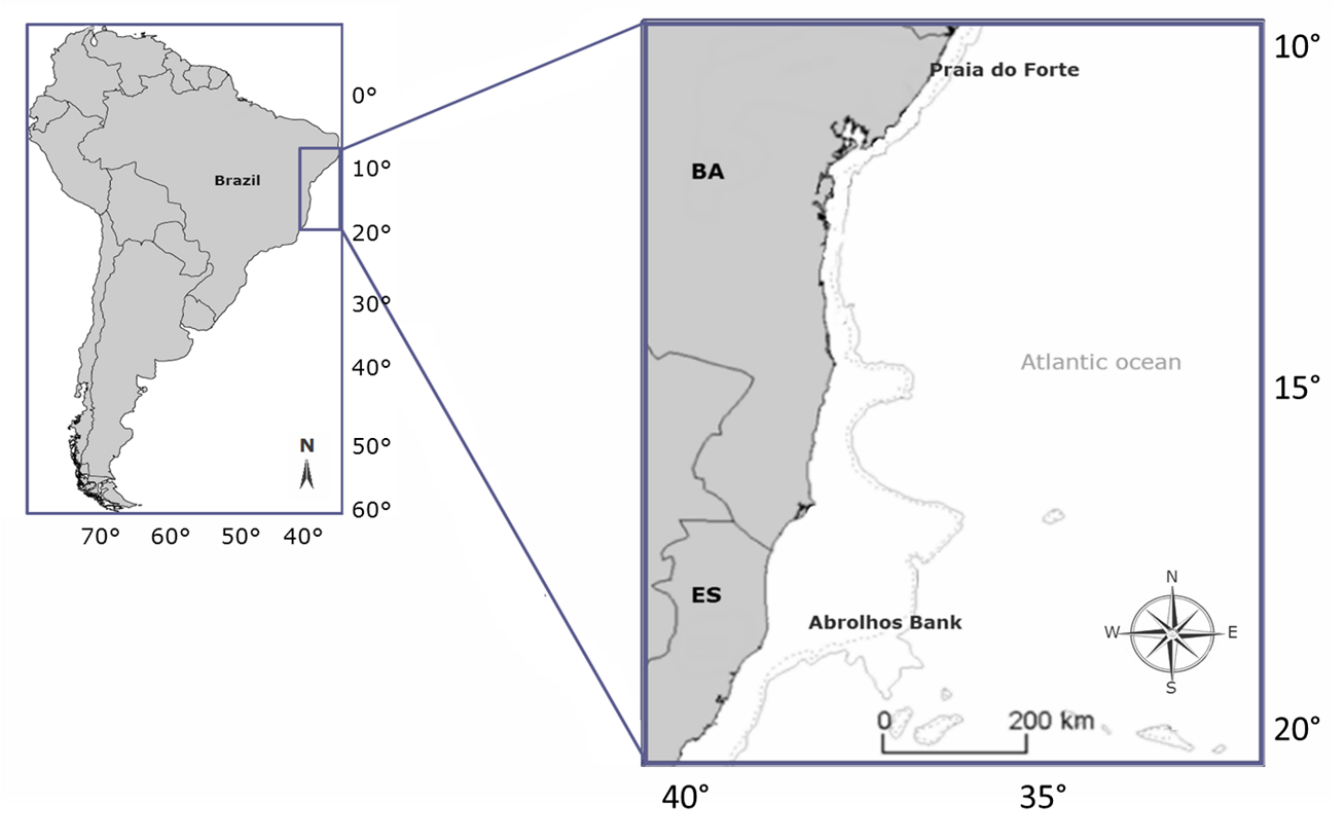
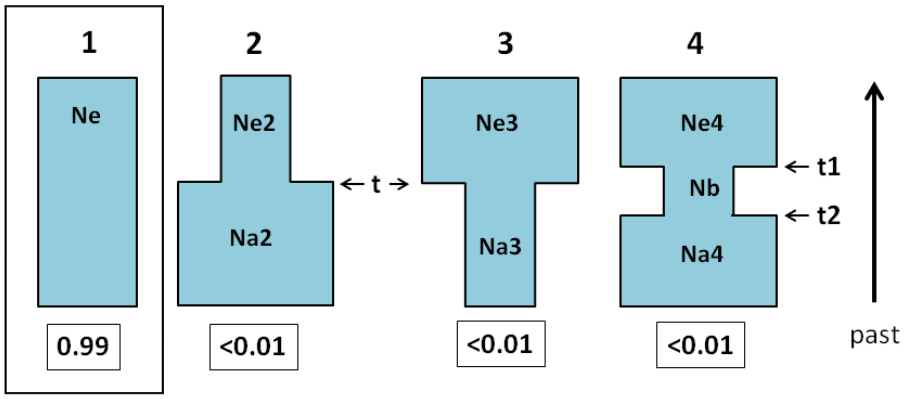
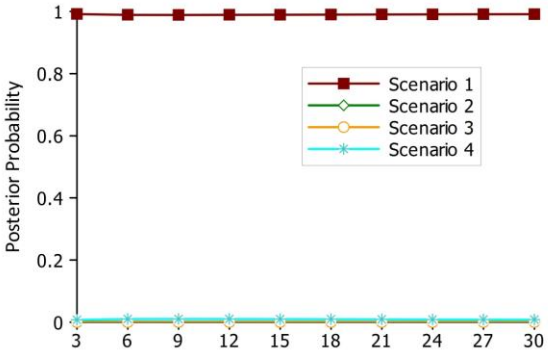

Brazil
N
0°
10°
20°
30°
40°
50°
60°
70°
60°
50°
40°
Praia do Forte
BA
ES
Atlantic ocean
Abrolhos Bank
N
W
E
S
0
200 km
10°
15°
20°
40°
35°

Figure 1

**a)**


1
2
3
4
Ne
Ne2
Na2
Ne3
Na3
Ne4
Nb
Na4
← t →
← t1
← t2
0.99
<0.01
<0.01
<0.01
past

**b)**


Posterior Probability
1
0.8
0.6
0.4
0.2
0
3
6
9
12
15
18
21
24
27
30
Scenario 1
Scenario 2
Scenario 3
Scenario 4

**c)**

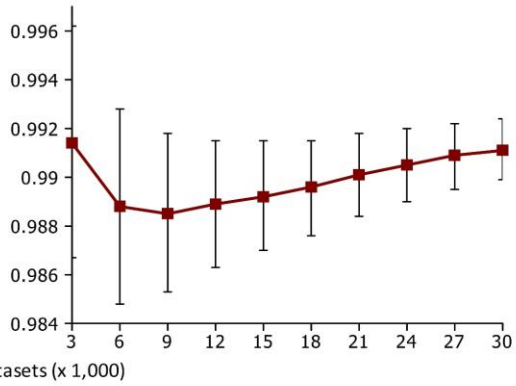

0.996
0.994
0.992
0.99
0.988
0.986
0.984
3
6
9
12
15
18
21
24
27
30
Closest simulated datasets (x 1,000)

Figure 2

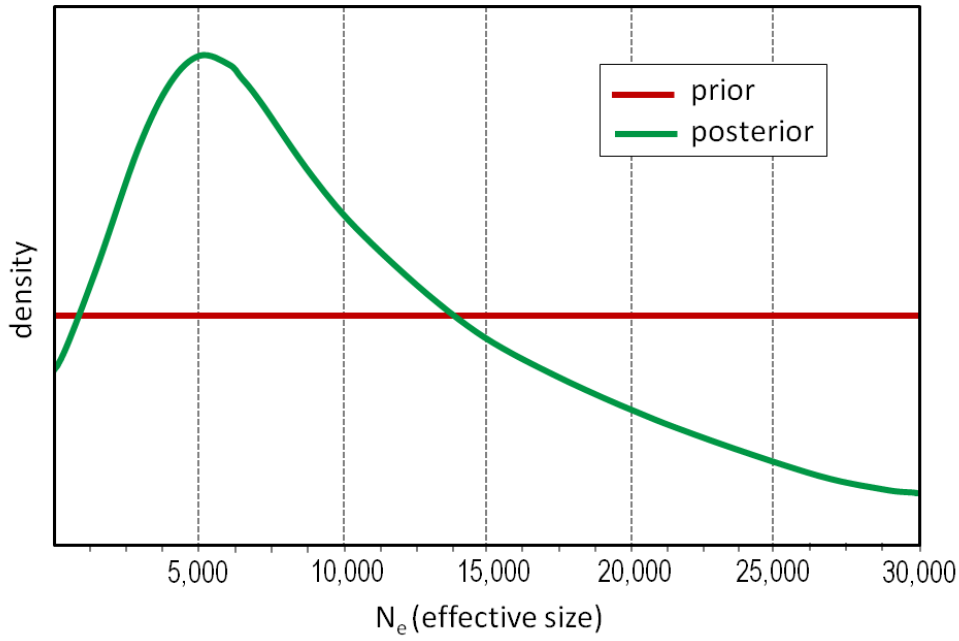

prior
posterior
density
5,000
10,000
15,000
20,000
25,000
30,000
Ne (effective size)

Figure 3

Table 1. Priors minimum and maximum (distribution uniform) for the parameters used for the four demographic scenarios (Fig. 2a) in the ABC approach. Effective sizes are in number of individuals and times are in number of generations (generation time of 18 years).

| Scenario<br>Parameter | Minimum | Maximum |
|---|---|---|
| Scenario 1 | | |
| *Ne* | 10 | 30,000 |
| Scenario 2 | | |
| *Ne2* | 10 | 300 |
| *Na2* | 5,000 | 30,000 |
| *t* | 2 | 10 |
| Scenario 3 | | |
| *Ne3* | 1,000 | 5,000 |
| *Na3* | 10 | 300 |
| *t* | 2 | 10 |
| Scenario 4 | | |
| *Ne4* | 1,000 | 5,000 |
| *Nb* | 10 | 300 |
| *Na4* | 5,000 | 30,000 |
| *t1** | 2 | 10 |
| *t2** | 2 | 10 |

* *t2>t1*

Table 2. Summary statistics for ten microsatellite loci genotyped for humpback whale population off Brazil. Rep, repeat motif length in base pairs; *K*, number of alleles; $H_O$, observed heterozygosity; $H_E$, expected heterozygosity; $F_{IS}$, inbreeding coefficient (*$P$<0.005 based on 180 randomizations).

| Locus | Rep | Allele range | *K* | $H_O$ | $H_E$ | $F_{IS}$ |
|---|---|---|---|---|---|---|
| GATA 28 | 4 | 143-203 | 15 | 0.626 | 0.612 | - 0.022 |
| GATA 53 | 4 | 231-287 | 14 | 0.791 | 0.835 | 0.053 |
| GATA 417 | 4 | 186-280 | 18 | 0.923 | 0.909 | - 0.016 |
| 199/200 | 2 | 102-118 | 8 | 0.567 | 0.549 | - 0.034 |
| 417/418 | 2 | 178-204 | 11 | 0.754 | 0.809 | 0.068 |
| 464/465 | 2 | 130-152 | 10 | 0.587 | 0.610 | 0.038 |
| EV1Pm | 2 | 121-129 | 5 | 0.553 | 0.532 | - 0.040 |
| EV37Mn | 2 | 192-224 | 17 | 0.900 | 0.923 | 0.026 |
| EV94Mn | 2 | 201-221 | 11 | 0.808 | 0.817 | 0.012 |
| EV96Mn | 2 | 183-215 | 17 | 0.854 | 0.866 | 0.014 |

Table 3. Effective ($N_e$) and census ($N_c$) population sizes estimates based on different methods and estimated in different time frames (generations).

| Method | $N_e$ (CI) | $N_c$ (CI) | Time frame (generations ago) |
|---|---|---|---|
| LDNe | 1,061 (688 - 2,104) | 3,820 (2,476 - 7,574) | 1 generation |
| ONESAMP | 560 (475 - 731) | 2,016 (1,710 - 2,632) | 2 to 8 generations |
| DIYABC | 4,170 (2,330 - 26,600) | 15,012 (8,388 - 95,760) | $4N_e$ generations |
| SMM* | 28,238 (5,648 - 56,476) | 101,656 (20,332 - 203,313) | $4N_e$ generations |

CI, 95% confidence or credibility interval.

* Mean and interval are estimated using 2 X $10^{-4}$, $10^{-4}$ and $10^{-3}$ mutation rates, respectively.

# Capítulo 3 - Genetic differentiation between humpback whales (*Megaptera novaeangliae*) from Atlantic and Pacific breeding grounds of South America

Ana Lúcia Cypriano-Souza, Márcia H. Engel, Susana Caballero, Lilian Flórez-González, Carlos Olavarría, Juan Capella, Debbie Steel, Angie Sremba, C. Scott Baker, Sandro L. Bonatto

**(Artigo a ser submetido à revista científica *Marine Ecology Progress Series*)**


“On one occasion I saw two of these monsters (whales) probably male and female, slowly swimming, one after the other, within less than a stone’s throw of the shore” (Terra Del Fuego), “over which the beech tree extended its branches.”

— Charles Darwin, *A naturalist's voyage around the world*

# Genetic differentiation between humpback whales (*Megaptera novaeangliae*) from Atlantic and Pacific breeding grounds of South America

Ana Lúcia Cypriano-Souza[1,2]; Márcia H. Engel[2]; Susana Caballero[3]; Lilian Flórez-González[4]; Carlos Olavarría[5]; Juan Capella[4]; Debbie Steel[6]; Angie Sremba[6]; C. Scott Baker[5,6] and Sandro L. Bonatto[1]

[1] *Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga 6681, 90619-900 Porto Alegre, RS, Brazil*

[2] *Projeto Baleia Jubarte/Instituto Baleia Jubarte, Rua Barão do Rio Branco 26, 45900-000 Caravelas, BA, Brazil*

[3] *Laboratorio de Ecología Molecular de Vertebrados Acuáticos-LEMVA, Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia*

[4] *Fundación Yubarta, Carrera 24F Oeste No. 3-110, Tejares de San Fernando, Cali, Colombia*

[5] *School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1010, New Zealand*

[6] *Marine Mammal Institute, Department of Fisheries and Wildlife, Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365, USA*

Address for correspondence: Ana Lúcia Cypriano-Souza, Faculdade de Biociências, PUCRS, Av. Ipiranga 6681, 90619-900 Porto Alegre, RS, Brazil, Phone: +55 (51) 3353.4727, Fax: +55 (51) 3320.3612, Email: anacypriano@yahoo.com.br

Running title: genetic differentiation between humpback whale breeding stocks A and G

**Abstract**

The humpback whale populations wintering off the eastern and western coast of South America are recognized as the Breeding Stocks A (BSA) and G (BSG), respectively. BSA humpbacks are distributed along the Brazilian coast while the whales from BSG are found mainly off Colombia and Ecuador coasts. Previous studies have provided the distinctiveness of these two breeding grounds of other Southern Hemisphere stocks, and the migratory connections with their respective feeding grounds. However, some issues regarding the degree of connectivity and differentiation between Southwestern Atlantic and Southeastern Pacific humpback whales still remain unknown. Here we present the first analysis of genetic differentiation and level of gene flow between BSA and BSG, based on genetic data from the mtDNA control region (465 bp) and 16 microsatellite loci from a total of 511 samples collected at the coasts off Brazil ($n$ = 277) and Colombia ($n$ = 148) breeding grounds, and around western Antarctic Peninsula feeding area ($n$ = 86). We corroborated previous studies that showed no genetic differentiation between the Colombia breeding ground and Antarctic Peninsula feeding area, supporting its migratory connection. Our results strongly supported population differentiation, at both mtDNA ($F_{ST}$ = 0.058) and microsatellite ($F_{ST}$ = 0.011) markers, between Breeding Stocks A and G, which was also evidenced by the Bayesian clustering analysis even without sampling location prior information. However, assignment tests suggest that exchange of individuals is occurring between two populations, but likely with a gene flow low enough to allow the demographic independence of these breeding grounds. Finally, our data segregated by gender showed a significant differentiation between females from Brazil and Colombia and a higher segregation between males from Brazil and Antarctic Peninsula. These findings suggest fidelity of females to the breeding areas and males loyalty to the feeding areas. However, for a better understanding of the connectivity between the breeding grounds a sampling effort should be performed mainly on arrival and departure of the whales migrating for these areas.

*Key words:* population structure, microsatellite, mtDNA, migration, individual assignment

## Introduction

Humpback whales (*Megaptera novaeangliae* Borowski, 1781) are widely distributed throughout the major ocean basins, where they undertake long-distance seasonal migrations between high latitude summer feeding grounds and low latitude breeding grounds (Clapham and Mead 1999, Dawbin 1966). Most humpback whale breeding and feeding grounds are well known and the populations show different degrees of fidelity to these areas. In the Northern Hemisphere (NH) the populations show high maternal fidelity to their feeding grounds, mixing on a single breeding ground in the North Atlantic and on common breeding grounds in the North Pacific Ocean (Baker et al. 1986, 1994, 1998a, Calambokidis et al. 2001, 2008, Palsbøll et al. 1995, 1997, Stevick et al. 2006a). In contrast to the segregated humpback whale feeding grounds in the Northern Hemisphere, the foraging areas of humpbacks in the Southern Hemisphere (SH) are distributed throughout a broad circumpolar area of the Southern Ocean where there is no continental barrier to dispersal, increasing the potential to movement among populations (Hoelzel 1998). Based on whaling records and biological data, these SH feeding areas were divided into six management areas, Areas I-VI, by International Whaling Commission (IWC) (Donovan 1991). In addition, based on wintering distribution of humpback whales in the SH seven breeding stocks (termed A-G) are recognized by IWC (IWC 2005). Humpback whales that spend year-round in the Arabian sea has been recognized as Breeding Stock X (IWC 2005, Mikhalev 1997).

The humpback whale breeding grounds off the eastern and western coast of South America are considered the Breeding Stocks A (BSA) and G (BSG), respectively (Fig. 1). In the Southwestern Atlantic Ocean, whales from BSA are distributed along the Brazilian coast from approximately 5° to 23° S (Andriolo et al. 2010, Zerbini et al. 2004), with additional sightings to the north and west of 5° S (Furtado-Neto et al. 1998, Magalhães et al. 2008, Meirelles et al. 2009, Pretto et al. 2009), and near oceanic islands as the Fernando de Noronha Archipelago and Trindade Island (Lodi 1994, Siciliano et al. 2012, Wedekin et al. in press). Current distribution data have demonstrated that the main mating and calving ground for this population is in the Abrolhos Bank (16°40'- 19°30' S and 37°25'- 39°45' W) (Andriolo et al. 2006, 2010, Freitas et al. 2004, Martins et al. 2001, Siciliano et al. 1997), and an increasing number of sightings of humpback whale groups, including females with calves, have been reported in the northern coast of Bahia, mainly off Salvador (Martins et al. 2001, Rossi-Santos et al. 2008, Zerbini et al. 2004).

In the Southeastern Pacific Ocean, humpback whales from BSG occur from northern Peru coast (6° S) to Costa Rica coast (12° N), with a very low density of whales found around

the Galápagos Archipelago, located 1,000 km west of the Ecuador coast (Acevedo-Gutiérrez and Smultea 1995, Félix and Haase 2005, Félix et al. 2006, 2009, Flórez-González et al. 1998, Pacheco et al. 2009). Although this breeding area comprises a wide range of approximately 3,000 km of coast, with at least five separate humpback whale calving areas, individual movement has been reported between Ecuador and Colombia, Colombia and Panama, Ecuador and Peru, Colombia and Peru, and Ecuador and Costa Rica, indicating high interchange of individuals among these areas (Castro et al. 2008, Félix et al. 2009, Flórez-González et al. 1998). Interestingly, BSG is considered one of Southern Hemisphere humpback whale populations, mating and calving during the austral winter, but the region has a large extension located in the Northern Hemisphere, off Central America (Acevedo-Gutiérrez and Smultea 1995, Flórez-González et al. 1998, Rasmussen et al. 2007). Furthermore, there is a spatial overlap with humpbacks from feeding areas off California that breed during the boreal winter in this area (Calambokidis et al. 2000). This spatial overlapping in a same breeding ground could favor trans-equatorial gene flow (mainly through males) between North and South Pacific humpback whales, as indicated by sharing of some genetic traits between these populations (Acevedo-Gutiérrez and Smultea 1995, Baker et al. 1993, Medrano-González et al. 2001, Olavarría et al. 2007, Rasmussen et al. 2007).

Recent studies involving different methods as photo-identification, satellite telemetry and genetic analyses have provided the distinctiveness of Breeding Stocks A and G of other SH humpback whale stocks and the migratory connections with their respective feeding grounds (Acevedo et al. 2007, Caballero et al. 2001, Cypriano-Souza et al. 2010, Engel et al. 2008, Engel and Martin 2009, Olavarría et al 2007, Stevick et al. 2004, 2006b, Stone et al. 1990, Zerbini et al. 2006, 2011). Migratory connections have been found between the feeding area around the western Antarctic Peninsula (AP) (~ 60° S, 64 W), belonging Antarctic Area I, and the Southeastern Pacific breeding areas, BSG. Comparisons through photo-identification catalogs from these areas have supported the link of AP with the wintering destinations off Colombia, Ecuador, Panama and Costa Rica (Rasmussen et al. 2007, Stevick et al. 2004, Stone et al. 1990). Additionally, mitochondrial DNA (mtDNA) analyses have demonstrated high similarity of the breeding grounds from Colombia and Ecuador with AP feeding area (Caballero et al. 2001, Félix et al. 2012, Olavarría et al. 2000). Although IWC established the stock boundary separating the feeding Areas I and II (western and eastern Antarctic Peninsula, respectively) at 60° W, this boundary was moved east to 50° W based on subsequent genetic and photo-identification data (Dalla Rosa et al. 2004, Engel et al. 2008, Olavarría et al. 2000). Recently, Dalla Rosa et al. (2008) have suggested that the boundary for

the Area I feeding area should be extended to the eastern of the Antarctic Peninsula, near South Orkney Islands (~ 40° W) in the Weddell Sea, as indicated by tracking of individuals in the Southern Ocean.

In contrast, the lack of photographic matches between whales from Antarctic Peninsula (western and eastern, Area I and II respectively) and Brazil (Stevick et al. 2004) and the significant mtDNA differentiation (Engel et al. 2008) between these two regions discard the hypothesis of a link of AP with BSA. Furthermore, Breeding Stock A have been linked to the feeding areas around South Georgia and South Sandwich Islands (between the latitudes 54° and 60° S, and the longitudes 33° and 22° W) in the Scotia Sea, belonging Antarctic Area II (Engel et al. 2008, Engel and Martin 2009, Stevick et al. 2006b, Zerbini et al. 2006, 2011). This migratory connection between Brazil and Area II in the Scotia Sea had previously been suggested by Slijper (1962, 1965) and Mackintosh (1965), but no direct evidence had been provided to support this hypothesis. Recently, the first evidence of the connection between BSA and South Georgia and South Sandwich Islands was by two humpbacks monitored by satellite telemetry (Zerbini et al. 2006), and further corroborated through individuals photo-identified in Abrolhos Bank and resighted near Shag Rocks (west of South Georgia) and near the South Sandwich Islands (Engel and Martin 2009, Stevick et al. 2006). In addition, two whales sampled in 2006 near island of South Georgia have presented mtDNA haplotypes identified in the Brazilian breeding ground (Engel et al 2008) of which one showed a putative parent-offspring relationship with one female sampled off Abrolhos Bank in 2001 (Cypriano-Souza et al. 2010).

The Magellan Strait (MS) located in the southwestern South America was also recognized as a feeding ground, where humpback whales have been observed feeding during the austral summer and fall (Acevedo et al. 2004, Gibbons et al. 2003). More recently, the migratory destinations for the humpbacks from Magellan Strait feeding area were identified through resightings of individuals photo-identified in Ecuador, Colombia, Panama and Costa Rica (Acevedo et al. 2007, Capella et al. 2008), indicating that these whales belong to BSG. Nevertheless, demographic (based on photo-identification) and genetic data have showed that this feeding aggregation is separate and genetically distinct from the Antarctic Peninsula feeding area, suggesting segregation of maternal lineages in these feeding areas of BSG, similar that observed for the humpback whale populations from the North Atlantic and North Pacific Oceans (Acevedo et al. 2007, Olavarría et al. 2006). Moreover, photo-identification data have demonstrated some heterogeneity in the breeding areas of BSG, as indicated by a higher interchange between MS and northernmost breeding areas (Panama and Costa Rica)

than between MS and southernmost breeding areas (Ecuador and Colombia) (Acevedo et al. 2007, Capella et al. 2008). Also, Félix et al. (2012) demonstrated a genetic heterogeneity between Ecuador and Colombia, suggesting that this difference could be explained for different migration patterns between sexes, in which females show higher site fidelity than males.

Despite previous studies have provided important information about the humpback whale breeding grounds from western and eastern coast of South America, some issues regarding the degree of connectivity and differentiation between these two populations still remain unknown. Interchange of individuals between SH breeding stocks has been showed through genetic analyses, acoustic data and photographic matches (Darling and Souza-Lima 2005, Noad et al. 2000, Pomilla and Rosenbaum 2005, Rosenbaum et al. 2009, Stevick et al. 2010, 2011). Based on mtDNA analyses of four SH breeding stocks (A, B, C and X) within the Southern Atlantic, the Southwestern and Northern Indian Oceans, Rosenbaum et al. (2009) have shown significant differentiation between these populations with a degree of gene flow, mainly between BSB and BSC with the highest numbers of migrants involving two different oceans. Further, a higher female structure than male structure was observed, suggesting female fidelity and increased male dispersal (Rosenbaum et al. 2009). In addition, genetic evidence of movement of a male whale has been reported from eastern to western Africa (BSC and BSB respectively) (Pomilla and Rosenbaum 2005). Interestingly, haplotypes of the clade AE, characteristic of humpback whales on the North Pacific and found only in BSG within the SH populations (Olavarría et al. 2000, 2007), have recently been reported in Brazil (BSA) (Engel et al. 2008). Current genetic study also suggested that recent gene flow through the three southern oceans has occurred, as indicated by haplotypes sharing of humpback whales of the BSG with those from distant breeding areas within Atlantic and Indian Oceans (Félix et al. 2012).

Darling and Souza-Lima (2005) observed song similarity between Gabon (BSB) and Brazil (BSA), suggesting that these stocks may interchange individuals and/or song on a common feeding ground or during the migration routes (Clark and Clapham 2004). Finally, two long movements between breeding grounds in two different ocean basins have been evidenced through photographic matches (Stevick et al. 2010, 2011). A female whale from Breeding Stock A travelled approximately 10,000 km to Madagascar, Breeding Stock C (Stevick et al. 2010). Another female accompanied by a calf sighted in Ecuador (BSG) was resighted travelling in a pair in Abrolhos Bank (BSA), representing the first record of a humpback whale in both the Pacific and Atlantic breeding grounds of South America (Stevick

et al. 2011). These findings demonstrate that humpback whale migrations are not exclusive to males, indicating a behavioral flexibility for an usually philopatric species (Stevick et al. 2010, 2011).

The present study evaluates the population structure and the level of gene flow between the humpback whales from breeding stocks G and A, western and eastern of South America respectively, based on the analysis of mtDNA control region haplotypes (465 bp) and microsatellite genotypes (16 loci). This report also presents a review of previous mtDNA control region data of Brazilian humpback whales (Engel et al. 2008, Rosenbaum et al. 2009). We also investigated the potential for sex-biased dispersal relative to population structure in the breeding grounds. Finally, this study presents the first analysis of population differentiation using microsatellite loci in humpback whales from BSG and BSA and the first use of genotypes to assess individual assignment and potential migrants in these breeding grounds.

## Materials and methods

*Sample collection, DNA extraction and sex determination*

A total of 511 skin samples of humpback whales, including 428 samples described in previous studies (Cypriano-Souza et al. 2010, Engel et al. 2008, Olavarría et al. 2000, 2007, Steel et al. 2008), were collected from two breeding grounds (BSA, $n$ = 277; and BSG, $n$ = 148) of South America during the winter breeding seasons (July-November), and in the Antarctic Peninsula feeding ground ($n$ = 86) during the austral summer (Fig. 1). Samples from BSA (Abrolhos Bank and Praia do Forte) were collected by Instituto Baleia Jubarte from 1997 to 2011, while samples from BSG (Gorgona Island and Colombia coast) were collected by Fundación Yubarta from 1991 to 1999. Several research groups collected samples from Antarctic Peninsula: Chilean Antarctic Institute (INACH) from 1996 to 1999, Southern Ocean Global Ecosystems Dynamics (SO-GLOBEC) in 2002, and IWC (3 samples used in this study) in 1990 and 1994. Most of the samples were obtained by the biopsy dart procedure (Lambertsen 1987) and some were collected from sloughed skin or stranded whales. Biopsy sampling of humpbacks from BSA followed the same guidelines performed in previous studies (Cypriano-Souza et al. 2010, Engel et al. 2008). Samples were preserved in 70% ethanol and were stored at -20°C prior to DNA extraction.

DNA had been previously extracted for 51 of the 134 Brazilian humpback samples (those collected between 1999 and 2005), which were used in the earlier studies (Cypriano-Souza et al. 2010, Engel et al. 2008). Total genomic DNA of the remaining samples ($n$ = 83,

collected between 2006 and 2011) was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) according to manufacturer's protocol.

Molecular sexing was carried out by PCR amplification followed by TaqI digestion of the genes ZFX and ZFY following the protocol of Palsbøll et al. (1992) modified by Bérubé and Palsbøll (1996). The sex ratio for each population was compared to an expected 1:1 ratio using the Pearson chi-square test with Yate's correction.

*Mitochondrial DNA control region sequencing and analyses*

Mitochondrial DNA control region sequences of 148 and 82 humpback samples from Colombia breeding ground (BSG) and Antarctic Peninsula feeding ground, respectively, used here have been described previously (Olavarría et al. 2000, 2007). Mitochondrial DNA diversity of the humpback whales breeding off the Brazilian coast (BSA) had previously been described in two publications (Engel et al. 2008, Rosenbaum et al. 2009). Engel et al. (2008) originally sequenced about 400 bp of the first segment of the mtDNA control region (using Dlp-1.5 and Dlp-5 primers) for 176 samples collected between 1997 and 2001. Rosenbaum et al. (2009) updated the previous publication and re-sequenced the mtDNA control region to a longer sequence length (486 bp) for some of these samples and identified 66 haplotypes from 64 polymorphic sites for 164 individuals. Of the 176 samples included in Engel et al. (2008), 12 (not re-sequenced) were not included in Rosenbaum et al. (2009) since they were described by a shorter sequence length (376 bp).

Here, a total of 59 samples previously analyzed (Engel et al. 2008, Rosenbaum et al. 2009), representing 59 unique haplotypes of the 66 described by Rosenbaum et al. (2009) were re-sequenced. Six haplotypes, which no DNA was available, were confirmed through reviewing of the electropherograms (of high quality) from the original sequencing for these samples. One haplotype (HBR014) represented by only one sample could not be confirmed since there is not DNA available or electropherogram with good quality. Reviewing of these haplotypes is important to allow a collaborative comparison of mtDNA control region haplotypes of humpback whales from BSG and Antarctic Peninsula feeding ground, and later comparison to historical haplotypes of humpback bones collected from South Georgia Island whaling stations, as described by Sremba et al. (2010).

For sequencing, a fragment of approximately 700 bp of the mtDNA control region was amplified using the primers M13Dlp1.5 (tPro whale, 5'-TGTAAAACGACGGCCAGTTCACCCAAAGCTGRARTTCTA-3') and Dlp8G (5'-GGAGTACTATGTCTGTAACCA-3'). PCR reactions contained: 2.5 mM $MgCl_2$, 0.2 mM

dNTPs, 0.4 μM of each primer, 0.125 U of Platinum Taq DNA Polymerase (Invitrogen), 1 X PCR buffer (Invitrogen), and 1 μl of DNA (approximately 20 ng). Thermocycle profile consisted of an initial denaturation at 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 40 s, and extension at 72°C for 40 s, and concluded with a final extension at 72°C for 10 min. PCR products were purified with ExoSap-IT (USB) before sequencing reaction with BigDye$^{TM}$ Dye Terminator Chemistry v3.1 (Applied Biosystems). Products for sequencing were cleaned using CleanSEQ (Agencourt) and run on an ABI 3730XL (Applied Biosystems) at Hatfield Marine Science Center (HMSC) of Oregon State University (OSU) in Newport, OR. Fifty-eight of the 59 samples provided clean sequence. The sequences were aligned and all variable positions were confirmed visually, according to the OSU electropherogram using SEQUENCHER 5.0 (Gene Codes Corporation). Sequences were trimmed to a consensus length of 464 bp and haplotypes were identified. Furthermore, for the validation of the 58 BSA haplotypes re-sequenced, a comparison to the GenBank sequences (AY329844–AY330096 for Engel et al. 2008, and GQ913691-GQ913852 for Rosenbaum et al. 2009) was carried out. As the first 22 bp were not clean for the OSU sequences, all sequences were trimmed to 464 bp.

Molecular diversity indices, such as haplotype ($h$) and nucleotide ($\pi$) diversity, and haplotype frequency were estimated for each population using the program ARLEQUIN v. 3.5 (Excoffier and Lische 2010). Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997) neutrality tests were calculated for each population with 1000 bootstrap replicates to evaluate demographic expansion or contraction in the same program. Genetic differentiation between populations was tested with pairwise F-statistics ($F_{ST}$, differences in haplotype frequency, and $\Phi_{ST}$, nucleotide differentiation) and analysis of molecular variance (AMOVA) by permutation procedure in ARLEQUIN (10 000 permutations, with significance set at $\alpha = 0.05$). The number of migrants per generation was estimated using the formula of Slatkin (1991) as described in ARLEQUIN.

*Microsatellite genotyping and analyses*

Samples were screened for genetic variation at 16 microsatellite loci (14 dinucleotides: EV1, EV14, EV21, EV37, EV94, EV96, EV104 (Valsecchi and Amos 1996), GT23, GT211, GT575 (Bérubé et al. 2000), 464/465 (Schlötterer et al. 1991), RW4-10, RW31, RW48 (Waldick et al. 1999), and two tetranucleotides: GATA28 and GATA417 (Palsbøll et al. 1997b)). Genotypes of 134 samples of the Brazilian humpback whales were assessed previously in Cypriano-Souza et al. (2010, in prep) for 7 of these 16 loci on a MegaBACE

1000 automated sequencer (GE Healthcare) at Laboratory for Genomic and Molecular Biology of Pontifical Catholic University of Rio Grande do Sul, in Brazil. Samples of the humpbacks from BSG ($n$ = 145) and Antarctic Peninsula ($n$ = 83) had been genotyped for all loci on an ABI 3730XL (Applied Biosystems, Foster City, CA), and these genotypes were provided by South Pacific Whale Research Consortium (SPWRC) (Steel et al. 2008) for a collaborative comparison between populations. To accomplish this comparison, 40 samples of the 134 samples from BSA were genotyped for all loci on an ABI 3730XL (Applied Biosystems) at HMSC of OSU in Newport, OR, allowing the calibration and standardization of microsatellite allele size.

PCRs were carried out in 10 μl with the following conditions: 1.5-4.0 mM of $MgCl_2$, 200 μM of dNTPs, 0.4 μM of each primer, BSA (Bovine Serum Albumin), 0.25 units of Platinum Taq DNA polymerase (Invitrogen), 1 X PCR buffer (Invitrogen), and 1 μl of DNA (~ 50 ng). All loci were amplified in separate reactions using the following thermocycle profile: initial denaturation at 94°C for 3 min, 35-40 cycles of denaturation at 94°C for 30 s, annealing at 50°C for 30 s, and extension at 72°C for 40 s, followed by a final extension at 72°C for 10-30 min depending on the locus. Seven samples with known genotypes were included as an internal control to ensure consistent allele sizing and to allow calibration of alleles. Negative controls were run at the PCR step to control for exogenous contamination. Amplicons were pooled in 4 sets of up to 5 loci and genotyped on an ABI 3730 and alleles were calling using GENEMAPPER v. 4.0 (Applied Biosystems).

The remaining of 94 samples of BSA humpbacks were amplified for the other 9 loci following exactly the same PCR conditions as described above. PCR products were checked on 1% agarose gel, visualized with GelRed™ (Biotium Inc.) and genotyped on an ABI 3730XL by Macrogen Inc. (Seoul, South Korea). Twelve samples with known genotypes were included to allow calibration of the alleles generated. Allele peaks were sized and visually verified using the free software Peak Scanner™ v1.0 (Applied Biosystems).

Analyses for matching genotypes (replicate samples and recaptures) were performed by GENECAP 1.4 (Wilberg and Dreher 2004) that compares each multilocus genotype with all others within the data set. Pairs of matching samples were subsequently compared for sex and mtDNA haplotypes data (when available) to support the identifying of replicates and/or recaptures. The probability of identity $P_{(ID)}$, the probability of two individuals sharing the same genotype by chance, was estimated with two different formulations (the Hardy-Weinberg [HW]$P_{(ID)}$, and the more conservative, Sib $P_{(ID)}$) using GENECAP. Duplicated samples were excluded for the further analyses. The program MICRO-CHECKER v.2.2.3

(Van Oosterhout et al. 2004) was used to check for possible null alleles, large allele dropout, and scoring errors due to stutter peaks.

Level of genetic diversity was estimated as the number of alleles per locus (*K*), observed and expected heterozygosities ($H_O$ and $H_E$, respectively) under Hardy-Weinberg assumptions (Nei 1978) for each locus in each population using CERVUS v. 3.0 (Kalinowski et al. 2007) and ARLEQUIN v. 3.5. The program FSTAT 2.9.3 (Goudet 2002) was used to calculate the Weir and Cockerham's (1984) measure of $F_{IS}$. Global and population-specific tests for deviations from Hardy-Weinberg equilibrium (HWE) (Guo and Thompson 1992) and linkage disequilibrium were carried out using the program ARLEQUIN, corrected for simultaneous comparisons with the sequential Bonferroni (Rice 1989).

Population genetic differentiation was tested with pairwise $F_{ST}$ measures (Weir and Cockerham 1984) and the related $R_{ST}$ index (Slatkin 1995), as well as with an AMOVA using ARLEQUIN. The statistical significance of $F_{ST}$ and $R_{ST}$ values was tested using 10 000 permutations. An unbiased estimate of differentiation ($D_{EST}$, Jost 2008) between the breeding stocks was also calculated using the software SMOGD 1.2.5 (Crawford 2010). The overall value of $D_{EST}$ for each pairwise population comparison was calculated as the arithmetic mean across loci, following Heller et al. (2010).

Population structure was also evaluated with a Bayesian clustering approach implemented in STRUCTURE v. 2.3.3 (Pritchard et al. 2000, 2007). Ten independent runs were performed for each *K* (number of cluster) between 1 and 5 with no prior information on sampling location using the admixture and correlated allele frequencies model. Burn-in and length of simulation were set at 500 000 and 1 000 000 iterations, respectively. In a second set of analyses the sampling location prior was used to assist the identification of clustering, as suggested for cases of subtle population structure (Hubisz et al. 2009). The results generated were processed in STRUCTURE HARVESTER v. 0.6.93 (Earl and vonHoldt 2012), a web-based program which plots the mean of the likelihood values per *K* and the ad hoc $\Delta K$ (Evano et al. 2005) to estimate the most likely value of *K*. The program also created input files for the analysis in CLUMPP (Jakobsson and Rosenberg 2007) which aligned the multiple results generated by STRUCTURE determining the optimal clustering which was graphically displayed by DISTRUCT (Rosenberg 2004). After identification of the populations number, the prior population information (assuming $K = 2$) was incorporated to assess individual assignment, possible migrants and individuals with admixed ancestry. Individuals with a membership coefficient (q) > 0.8 were considered residents for the population from which they were sampled, while those with $0.2 < q > 0.8$ were considered to be potentially admixed

(neither resident nor migrant) (Bergl and Vigilant 2007, Lecis et al. 2006). The "Use PopInfo" option (G = 0) was applied, and the length of simulation and burn-in were the same as previously described.

Additionally, another Bayesian method (Rannala and Mountain 1997) was performed to assess assignment/exclusion of individuals with predefined populations using GENECLASS v. 2.2.2 (Piry et al. 2004), which does not assume that all potential source populations have been sampled. The probability that an individual was a resident was calculated using a Monte Carlo resampling algorithm with 10 000 simulated individuals and an alpha of 0.1 (Paetkau et al. 2004). Each individual was assigned to the population in which the likelihood of its genotype is the highest, comparing the likelihood of the population where the individual was sampled with those of the available populations ($L = L_{home}/L_{max}$).

Sex-biased dispersal was tested using the program FSTAT v. 2.9.3 (Goudet et al. 2002) that calculates $F_{ST}$, inbreeding coefficient ($F_{IS}$), mean corrected assignment index (*mAIc*) and variance of corrected assignment index (*vAIc*) based on sex-specific expectations. The two-tailed test was used with 10 000 permutations. While $F_{ST}$ and *mAIc* are expected to be higher in the more philopatric sex, $F_{IS}$ and *vAIc* should be lower (Goudet et al. 2002). In addition, sex-specific estimates of genetic differentiation were calculated for both mtDNA and microsatellite data and compared to the data obtained in FSTAT.

## Results

*Validation of mtDNA control region haplotypes*

The comparison between the 58 BSA re-sequenced haplotypes and the GenBank sequences of previous studies (Engel et al. 2008, Rosenbaum et al. 2009) initially validated 47 haplotypes. However, six of these 47 collapsed into three unique GenBank haplotypes (most of sequences with length of 486 bp) due the slightly shorter consensus region (464 bp) considered in this study. Thus, we validated 44 haplotypes, but the six haplotypes need to be re-sequencing with the primers (Dlp-1.5 and Dlp-5, Baker et al. 1993) used in the original study for the validation of the other three haplotypes. Of the remaining 11 haplotypes (singletons), one showed a heteroplasmy and was not included in the next analyses, and 10 were not validated since they revealed misidentification at one to 5 variable sites. After review and correction, each of these 10 sequences collapsed into other known GenBank haplotypes, including the haplotype for which the single sample available did not have a clean sequence of the 59 provided for the re-sequencing.

In addition, six BSA haplotypes published in GenBank, which no DNA was available, were confirmed through reviewing of the electropherograms (of high quality) from the original sequencing for these samples. Based on this, a total of 51 haplotypes were confirmed of the 66 BSA haplotypes originally described in earlier studies (Engel et al. 2008, Rosenbaum et al. 2009). Also, given that three haplotypes described in the Rosenbaum et al. (2009) for the breeding grounds of South Atlantic and Indian Oceans (not included in BSA) were also found in six humpback individuals (one haplotype found in 4 samples, and the other two found in only one sample each) from Brazil, in this study we added them to the mtDNA diversity of the humpbacks from BSA totaling 54 haplotypes.

*mtDNA diversity*

For the alignment of the mtDNA control region sequences (464 bp) of 158 Brazilian humpback whale samples, 54 haplotypes were defined by 60 variable sites, including 2 insertions/deletions, 2 transversions and 56 transitions (Figure S1). After removal of replicate samples ($n$ = 18) based on genotype identity, the alignment of the mtDNA control region sequences for 130 and 82 humpback samples from Colombia breeding ground (BSG) and Antarctic Peninsula feeding area, respectively, resulted in 41 (2 transversions and 3 insertions/deletions) and 40 (3 transversions and 2 insertions/deletions) variable sites defining 27 and 21 haplotypes, respectively. When all 370 samples of the three populations were combined, a 465 bp consensus sequence revealed 77 haplotypes defined by 67 variable sites, including 3 insertions/deletions, 3 transversions and 61 transitions. Of these 77 haplotypes, 18 were shared between Colombia and Antarctic Peninsula, seven between Brazil and Colombia, and four between Brazil and Antarctic Peninsula, the same four shared among all populations. The frequencies of haplotypes for each population is shown in the Figure S1.

Overall haplotype diversity ($h$) was 0.962 (SD = 0.005) and nucleotide diversity ($\pi$) was 0.019 (SD = 0.00032) (Table 1). Although all three populations showed higher haplotype diversity, the Brazilian humpbacks presented a slightly higher $h$ (0.973, SD = 0.004) than whales from Colombia ($h$ = 0.906, SD = 0.015) and Antarctic Peninsula ($h$ = 0.902, SD = 0.020). In addition, the nucleotide diversity was higher for whales from Brazil ($\pi$ = 0.020, SD = 0.00047) compared to those from Colombia and Antarctic Peninsula, that presented identical values ($\pi$ = 0.018, BSG SD = 0.00052, and AP SD = 0.00095) (Table 1). Neutrality tests were non-significant and near zero for BSG and Antarctic Peninsula while Fu's Fs was significantly negative (-17.470, $P$ = 0.003) for BSA.

*Microsatellite diversity*

Three hundred and twenty seven samples (BSA, *n* = 130 from 1999-2011; BSG, *n* = 124; and Antarctic Peninsula, *n* = 73 from 1990-1999) were genotyped for at least 11 microsatellite loci, of which 313 (95,7%) were genotyped for all 16 loci. The loci were adequate for individual identification since that the probability of identity calculations were very small (HW $P_{(ID)} = 2.01 \times 10^{-18}$, and Sib $P_{(ID)} = 4.65 \times 10^{-7}$), indicating that even related individuals would have a low probability of sharing identical genotypes. Based on genotype identity and accessory information such as sex and mtDNA haplotype matches (when available), 13 genotypes were recaptures, of which 12 were recaptured along the Colombia coast, and one previously reported match between BSG and the Antarctic Peninsula (Steel et al. 2008). No match was found between the two breeding grounds (BSA and BSG). Thus, genotypes were assigned to 314 different individuals that were included in the further analyses.

No evidence of null alleles was detected, except for the loci 464/465 ($P < 0.01$) and EV94 ($P < 0.05$) in the BSG. No evidence for stuttering or large allele dropout was seen at loci. Deviation from Hardy-Weinberg equilibrium (HWE) was detected at the EV96 locus in BSG after Bonferroni correction, but was retained in the analyses since its removal did not affect the tested results. All pairwise comparisons of allele frequencies showed no significant linkage disequilibrium after Bonferroni correction.

All loci were polymorphic in all populations (Table 2), with the number of alleles per locus for all samples pooled ranging from 4 (EV1) to 20 (EV37) with a mean of 10.8. The mean observed and expected heterozigosities for the total sample were 0.72 and 0.74, respectively. For each population analyzed separately (Table 2) the number of alleles per locus ranged from 4 (EV1 for all populations) to 17 (EV37 and GATA417 for BSA), the observed heterozigosity ranged from 0.32 (GATA28 for Antarctic Peninsula) to 0.96 (GATA417 for BSA), and the expected heterozigosity from 0.32 (GATA28 for Antarctic Peninsula) to 0.93 (EV37 for BSA). FIS values were low for all loci (below 0.19) and were not significant for each population separated and for the total sample (Table 2).

*Population structure*

Bayesian clustering analyses with STRUCTURE clearly indicated a population structure with the highest posterior likelihood [mean Ln P(D) = -16402.0] for $K = 2$ clusters (Fig. 2) when compared to that for other numbers of clusters for all ten replicate runs, and this is also corroborated by the $\Delta K$ method of Evano et al. (2005) (Figure S2). The two clusters

correspond to the individuals from BSA and from BSG + Antarctic Peninsula (Fig. 2). There is no evidence for any differentiation between BSG and Antarctic Peninsula samples as seen in the $K$ = 3 bar plot graphics (Fig. 2). The results were similar using or not using the sampling location prior, although as expected in population with weak structure, in the latter the distinction between the clusters was not so clear.

Overall, the pairwise F-statics and AMOVA results corroborated the Structure results. Pairwise $F_{ST}$ comparisons corroborated that BSG and Antarctic Peninsula could not be distinguished and the significant differentiation between either of the two and BSA (Table 4). The AMOVA based on mtDNA data showed significant overall differentiation among the three sampling sites ($F_{ST}$ = 0.044; $\Phi_{ST}$ = 0.041) (Table 4). In addition, comparing BSA with BSG combined with Antarctic Peninsula resulted in significant highest overall differentiation at the haplotype and nucleotide level ($F_{ST}$ = 0.058; $\Phi_{ST}$ = 0.053) (Table 3). Similar results were also found in microsatellite allele frequencies given that the pairwise $F_{ST}$ comparisons were significant between BSA and both BSG and AP. But pairwise $R_{ST}$ was significant only between BSA and BSG (Table 4). Moreover, significant overall $F_{ST}$ (0.008) and high $D_{EST}$ (0.018) among the three populations, but no significant differentiation for overall $R_{ST}$ (0.005). When BSG was pooled with Antarctic Peninsula to form a single population, and subsequently compared to the BSA significant overall microsatellite differentiation was found for all F-statics ($F_{ST}$ = 0.011, $R_{ST}$ = 0.008, $D_{EST}$ = 0.024) (Table 3).

*Sex determination and sex-biased dispersal*

Sex was identified for 118 (82 males and 36 females) individuals from Colombia, for 86 (40 males and 46 females) individuals from Antarctic Peninsula, and for 264 (141 males and 123 females) individuals from Brazil. The sex ratio of the Colombian humpbacks (2.27:1 for males) was significantly skewed toward males ($X^2$ = 17.93, $P < 0.05$) but the sex ratios of the humpbacks sampled in Brazil (1.14:1 for males) and Antarctic Peninsula (1.15:1 for females) did not differ significantly from parity (respectively: $X^2$ = 1.22, $X^2$ = 0.42, $P > 0.05$).

For the sex-biased dispersal analyses based on microsatellites data, we used 127 samples (67 males and 60 females) from Brazil, 90 samples (66 males and 24 females) from Colombia, and 65 samples (25 males and 40 females) from Antarctic Peninsula. For the analysis based on mtDNA data we used 152 samples (83 males and 69 females) from Brazil, 115 samples (79 males and 36 females) from Colombia, and 82 samples (38 males and 44 females) from Antarctic Peninsula. It should be noted that for Brazil the sets of samples for the different markers consist mostly of different individuals. FSTAT tests showed no

significant evidence for sex-biased dispersal between BSA and BSG. Although females demonstrated higher *mAIc* and lower *vAIc* than males, suggesting male-biased dispersal and philopatry among females, these differences were not significant (Table 5). Moreover, the males showed a negative $F_{IS}$ which should be expected for the philopatric sex.

Sex-specific estimates of differentiation using mtDNA haplotype frequencies ($F_{ST}$) and molecular distances ($\Phi_{ST}$) (Table 3) were highly significant for both males and females but the overall $\Phi_{ST}$ value estimated for females ($F_{ST}$ = 0.070) was more than twice higher than that estimated for males ($\Phi_{ST}$ = 0.028). Interestingly, in the pairwise comparisons for females the differentiation between BSA and BSG was significant and twice higher than that between BSA and AP, while in the comparisons for males the $F_{ST}$ and $\Phi_{ST}$ values were significant and higher between BSA and Antarctic Peninsula (Table 6). The pairwise comparisons between BSG and Antarctic for females and males were low and not significant. When the latter two were pooled to compare with the BSA significant overall mtDNA differentiation was found for females ($F_{ST}$ = 0.087; $\Phi_{ST}$ = 0.085, $P$ = 0.000) and males ($F_{ST}$ = 0.052; $\Phi_{ST}$ = 0.033, $P$ = 0.000). On the other hand, while the sex-specific estimates of $F_{ST}$ base on microsatellites were significant (except among the three regions separated) and similar for both females and males, the $R_{ST}$ estimates were not significant and lower for both (Table 3).

*Migration rates and identification of migrants*

Incorporating local sampling as prior population information (assuming $K$ = 2) for STRUCTURE analyses, most (96.2%) individuals had high probability of being residents (q > 0.8) for the population in which they were sampled. However, six individuals, that did not assign clearly to the region from which they were sampled, were considered as potential migrants or with admixed ancestry (q < 0.8) (Figure 3), suggesting a low gene flow between the breeding grounds. Furthermore, these individuals were also assigned to the other population relative to their local sampling by the assignment analysis in GENECLASS. Four first-generation (F1) migrants were identified in the GENECLASS analysis with a probability below the threshold (P < 0.01). Three of these individuals (BR386, BR670 and Mno92Co020) were also identified as potential migrants (q > 0.9) between the regions by STRUCTURE (Table 7), two of them were sampled in Brazil and considered migrants from Colombia and one sampled in Colombia was a migrant from Brazil.

Based on Slatkin formula, the number of migrants per generation (*Nm*) estimated between BSA and BSG pooled with Antarctic Peninsula as a single population was 57.9 (based on $R_{ST}$) or 43.8 (based on $F_{ST}$). However, the number of female migrants per

generation was 8.9 (based on $\Phi_{ST}$) or 7.9 (based on $F_{ST}$), suggesting a lower gene flow between females. These estimates suggest an interchange of around 3 individuals per year and less than 1 female every two years between the breeding grounds (assuming a generation time of around 18 years, Roman and Palumbi 2003).

## Discussion

*Validation of mtDNA control region haplotypes*

Recently, forensic phylogenetic analyses of published mtDNA sequences have showed that error free publications are extremely rare (Bandelt et al. 2001, 2002). One of the errors most common is the called "phantom mutation" that may be generated in the sequencing (e.g. biochemical problem with the sequencing kit), misreading of the outputs or manual data transcription (Bandelt et al. 2001). The erroneous polymorphisms (e.g. false transversions, insertions and deletions) should show a mutation pattern that significantly differ from that of natural mutations (Bandelt et al. 2002). Indeed, the ten mtDNA control region haplotypes that were not validated showed misidentification at one to 5 polymorphic sites (mainly transversions), probably due misreading of the electropherograms with low quality. As these haplotypes were singletons, they have been more favor to error since they could not be confirmed by another sample with the same sequence. However, although these "singleton" errors did not affect significantly the previous results at population level, correct sequences are essentials to comparisons based on shared haplotypes, such as those with historical samples of this species (e.g. Shremba et al. 2010). Therefore, the reviewed sequences presented here are an important update to the genetic diversity of the Brazilian humpback whales, and will allow more reliable comparisons with data from other humpback whale populations.

*Population structure and breeding stocks identification*

Our results based on mtDNA and microsatellites genotypes, in special the results of the Bayesian clustering analyses with STRUCTURE (Fig. 2), strongly supported the previous mitochondrial evidences of no genetic differentiation between the Colombia breeding ground and Antarctic Peninsula feeding area (Caballero et al. 2001, Engel et al. 2008, Félix et al. 2012), as it has been also indicated by photographic and genotype matches between these areas (Steel et al. 2008, Stevick et al. 2004, Stone et al. 1990). Based on these evidences the humpback whales from Colombia and Antarctic Peninsula were pooled to form a single population representing the breeding stock G. Our results also strongly support that the

individuals studied here are structured in two genetically different populations that correspond to the western and eastern breeding grounds of South America, supporting the recognition of these two Southern Hemisphere stocks, respectively BSG and BSA. Overall, our tests for population differentiation based on both mtDNA ($F_{ST}$ = 0.058) and microsatellite ($F_{ST}$ = 0.011) markers revealed low but significant differentiation between BSA and BSG. This significant differentiation between BSA and BSG was also evidenced by the Bayesian clustering analysis done without sampling location prior information (Fig. 2), which still was able to detect the two populations. This is relevant since studies between geographically close whale populations using STRUCTURE analyses without using the sampling location prior usually could not detect the putative population structure (e.g. Carroll et al. 2010, Schmitt et al. 2012, Sremba et al. 2012).

Strong population structure have been reported among the feeding areas of the North Pacific ($F_{ST}$ ~ 0.5) and North Atlantic ($K_{ST}$ ~ 0.04) Oceans, as well as among the breeding areas in the North Pacific ($\Phi_{ST}$ ~ 0.3 for mtDNA haplotypes, and $\Phi_{ST}$ ~ 0.1 for nuclear intron alleles) (Baker et al. 1998a, Palsbøll et al. 1995). Nonetheless, in the Southern Hemisphere a weaker population differentiation have been reported among the breeding grounds, at both the Atlantic ($F_{ST}$ ~ 0.005 - 0.016) and Pacific Oceans ($F_{ST}$ ~ 0.014 - 0.079) based on the mtDNA analysis (Olavarría et al. 2007, Rosenbaum et al. 2009). Although our analysis have shown a low level of genetic differentiation between BSA and BSG stocks at both mtDNA and microsatellite, the $F_{ST}$ values were one of the highest of those estimated among other SH populations. However, it is important to highlight that BSG have also shown the highest differentiation with all South Pacific breeding areas ($F_{ST}$ ~ 0.058 - 0.079), particularly with those geographically closest, French Polynesia and Cook Islands, recognized as Breeding Stock F, and with the migratory corridor off eastern Australia (Olavarría et al. 2007, Valsecchi et al. 2010).

Now, the significant genetic differentiation between BSG and BSA, in the Southwestern Atlantic, shown in this study supports the previously suggested isolation of the former breeding ground. Indeed, six haplotypes (SP32, SP60, SP61, SP90, SP98 and SP101) shared among the breeding stock G, Antarctic Peninsula and Magellan Strait feeding areas have not been found in other SH populations (Félix et al. 2012). Besides, BSG together with its corresponding feeding grounds (Antarctic Peninsula and Magellan Strait) represent the only populations of the Southern Hemisphere where the haplotypes of the clade SH, characteristic of this hemisphere, were not found. This higher differentiation of BSG from the other SH populations could be due to an influence of historical or current trans-equatorial

gene flow taking place in the overlapping breeding ground between the humpbacks of the two hemispheres, that occurs off Central America (Acevedo-Gutiérrez and Smultea 1995, Baker et al. 1993, Medrano-González et al. 2001, Olavarría et al. 2007, Rasmussen et al. 2007). Despite this differentiation of the Colombian breeding ground, gene flow has been suggested between BSG and other breeding grounds from Oceania (BSE and BSF) and Western Australia (BSD), although at a very low degree, with interchange estimated to less than one female every two years (Olavarría et al. 2007).

*Genetic diversity*

Our data confirm a high genetic diversity for both mtDNA (Engel et al. 2008, Olavarría et al. 2007) and microsatellite (Cypriano-Souza et al. 2010) markers in the humpback whale populations that winter along western (microsatellites data for the first time for BSG) and eastern coasts of South America, in agreement with other breeding grounds in the Southern Hemisphere (Félix et al. 2012, Garrigue et al. 2004, Olavarría et al. 2007, Pomilla and Rosenbaum 2006, Rosenbaum et al. 2009, Valsecchi et al. 2002). The commercial whaling during the 20th century reduced the humpback whale populations to small percentages of their pre-exploitation sizes in the Southern Hemisphere, where about 200,000 humpbacks were caught (Gambell 1973, Tønnessen and Johnsen 1982). Despite this overexploitation the high genetic diversity found for these SH populations is likely due to the relatively brief duration of the population bottleneck, about six decades, which correspond to three to six generations (generation time between 12 and 24 years, Chittleborough 1965, Roman and Palumbi 2003) for the species as well as to a not too small minimum absolute population size (Amos 1996, Engel et al. 2008).

The nuclear and mitochondrial genetic diversity of the whales from Colombia (separately or together with those from Antarctic Peninsula) were slightly lower than those from Brazil, corroborating the previous results in which BSG has shown the lowest mtDNA diversity among the SH humpbacks populations (Félix et al. 2012, Olavarría et al. 2007, Rosenbaum et al. 2009, Valsecchi et al. 2010). The possible causes for BSG lower genetic diversity are not known. The simpler explanation is that this population size has been historically lower than the others, but unfortunately there is no estimate for the historical abundance of this population. A lower genetic variability could also have been attained by a relatively recent recolonization of the Southeastern Pacific Ocean after the glacial and interglacial oscillations of the climatic conditions of the Eastern Pacific (Baker et al. 1993, Lambeck et al. 2002). However, the neutrality tests have not indicated a recent expansion for

the whales from BSG, as should be characteristic of recent demographic events like recolonization followed by a population expansion. Yet another possibility is that this population suffered a population bottleneck by whaling much stronger that the BSA. However, this is unlikely since modern whaling in both the west and east coasts of South America were insignificant compared to other continents (Tønnessen and Johnsen 1982). Besides, although there is no estimate for the intensity of the exploitation bottleneck in BSG population, humpback whales have been seen in the Southeastern Pacific Ocean throughout their history, even in times of the lowest abundance (Félix and Haase 2005, Ramírez 1988).

*Migration rates, identification of migrants and sex-biased dispersal*

Notwithstanding the clear difference between the two breeding areas of South America shown here, the level of genetic divergence is relatively low which indicates a recent divergence and/or a relatively high (historical and/or ongoing) gene flow between these populations. Actually, recent gene flow among the three southern ocean basins have been suggested based on haplotype sharing between BSG and other breeding areas from all southern oceans, even with those more distant areas, within Atlantic and Indian Oceans (Félix et al. 2012). Although these shared haplotypes could be the result of shared ancestry, several movements across humpback whale stocks have been reported in the Southern Hemisphere (Pomilla and Rosenbaum 2005, Robbins et al. 2008, Steel et al. 2008, Stevick et al. 2010, 2011).

Our previous study has showed mtDNA haplotype sharing between BSA and both BSG and Antarctic Peninsula. Engel et al. (2008) reported two shared haplotypes between BSA and the Areas I and II around the Antarctic Peninsula, and comparing haplotypes from Brazilian humpback whales with those from Colombian breeding area (BSG), described by Olavarría et al. (2007), seven were shared between these two populations. Importantly, the clade AE has been also found in the Brazilian humpback whales (Engel et al. 2008), suggesting an interchange from BSG and/or Antarctic Peninsula to BSA. Since the samples from Brazil and Colombia used here are the same from the previous studies (Engel et al. 2008, Olavarría et al. 2007), our analyses identified the same seven shared haplotypes between these two stocks and four among the three sampling sites. More recently, an interchange of a female from Ecuador (BSG) to Abrolhos Bank (BSA) was reported, which was the first evidence of a migrant between these two breeding areas (Stevick et al. 2011).

However, for a better understanding of the differentiation and the level of connectivity between BSA and BSG, estimates of migration parameters were performed. The two main

approaches that have been used to estimate migration parameters are the coalescent and multilocus genotype (assignment) approaches (Faubet et al. 2007, Palsbøll et al. 2010, Waples and Gaggiotti 2006). While the coalescent methods estimate long-term evolutionary parameters, the assignment methods estimate short-term ecological parameters (Faubet et al. 2007). However, for reliable estimates of migration the levels of differentiation need to be intermediate to high ($F_{ST}$ > 0.05 to 0.10) (Faubet et al. 2007, Palsbøll et al. 2010). Consequently, when the genetic divergence among populations is low, the population-based methods should be complemented with kinship-based methods (Palsbøll et al. 2010). Our data have shown low to moderated estimates ($F_{ST}$ = 0.011 for microsatellites, and $F_{ST}$ = 0.058 for mtDNA) of genetic differentiation between the breeding stocks, allowing the estimate of migration parameters.

Most individuals were assigned to the source population when prior information was included in the Bayesian analysis, suggesting that current migration is low between the populations. On the other hand, six individuals were considered as putative migrants or admixed individuals of which three were also identified as first-generation migrants in the GENECLASS analysis, indicating some recent connectivity between these populations. Additional support for this evidence derives from the interchange of a female from Ecuador to Abrolhos Bank (Stevick et al. 2011). Importantly, there are two kinds of migration: the gene flow with transfer of gene copies among populations even without permanent displacement of individuals; and the dispersal which there is permanent dislocation of individuals among populations but without movement of gene copies into the recipient population (Palsbøll et al. 2010). Nonetheless, the identification of what kind of migration is estimated through genetic data is difficult. Our assignment tests suggest that in the migration between BSG and BSA may be occurring gene flow as has been indicated by the admixed individuals, which support an exchange of gene copies between the populations. Furthermore, the estimates of approximately 51 migrants and/or 8 females per generation, corresponding at 3 migrants per year and less than 1 female every two years respectively, indicate that exchange of individuals is occurring between the populations, but likely not enough to prevent their genetic differentiation.

The social organization of humpback whales have been characterized by a higher degree of female fidelity in both breeding and feeding grounds and a male tendency to dispersal (Baker et al. 1998b, Félix et al. 2012, Rosenbaum et al. 2009, Weinrich et al. 2006). In support of the expectation of female philopatry and male dispersal, reproductive pattern commonly found in several mammal species (Greenwood 1983), the genetic differentiation

between the humpback whales from BSG and BSA was stronger for mtDNA than microsatellite loci. One factor that may contribute to this reduced level of nuclear differentiation compared to mtDNA is its larger effective population size (Baker et al. 1998b, Palumbi and Baker 1994) but on the other hand its faster mutation rates should compensate it. Nonetheless, a lower gene flow between females have also been indicated by the migrant estimates of less than 1 female every two years compared to the 3 migrants (male and female together) per year. In addition, mtDNA differentiation estimates were significant for both females and males but the divergence between females was more than twice higher than that between males. Further, although similar $F_{ST}$ estimates based on microsatellites were significant for females and males, the $R_{ST}$ estimates were lower in general and did not show significant differences between sexes, suggesting again that males disperse more widely than females between the breeding grounds.

Although our FSTAT tests were not significant, sex-biased dispersal cannot be ruled out since these tests based on microsatellites have not capacity to detect sex-bias when the dispersal is low and the bias is small (Goudet et al. 2002), as it has been indicated by our results of significantly low nuclear differentiation. This finding is in agreement with previous studies for other cetacean species in which the tests of sex-biased dispersal were not significant even with a higher mtDNA differentiation than that of the microsatellites (Carroll et al. 2011, Hammer et al. 2012, Hollatz et al. 2011). However, genetic differentiation by sex may also be related to variability in migratory patterns of the whales. Temporal differences in occupancy pattern of breeding areas between individuals arriving from different feeding areas have been described in the North Atlantic (Stevick et al. 2003). Another behavioral pattern that may introduce heterogeneity is the migration of males more often than females to winter grounds (Brown et al. 1995, Craig and Herman 1997, Craig et al. 2003). In addition, males sharing the same feeding area may occasionally migrate to different breeding grounds in different wintering seasons (Darling and Cerchio 1993, Garrigue et al. 2002, Constantine et al. 2007). Valsecchi et al. (2010) have shown that the humpback whale migratory pattern seems to be more complex than previously thought, suggesting that males to maximize mating opportunities, should undertake longer longitudinal movements than females that need to save energy to reproduction. In contrast, long movements between breeding grounds in different oceans have been reported for two females, indicating a behavioral flexibility for the species (Stevick et al. 2010, 2011).

Recently, Félix et al. (2012) have demonstrated stratification between adjacent calving areas in the Southeastern Pacific Ocean, suggesting different migration patterns between

sexes, in which females show higher site fidelity than males. Interestingly, our data showed a significant higher differentiation for females between Brazil and Colombia but a higher segregation for males between Brazil and Antarctic Peninsula, suggesting that females show strong fidelity to their breeding areas and that males may be loyal to their feeding areas but more open to change the wintering destinations. Therefore, if gene flow is occurring between the two breeding grounds, it is unknown where during seasonal migration this could be happening. Given the distinctiveness of the two feeding areas (Antarctic Peninsula and Magellan Strait) of the Breeding Stock G, heterogeneity between the wintering areas of this stock and higher similarity between Magellan Strait and northernmost breeding areas (Acevedo et al. 2007, Capella et al. 2008), an alternate hypothesis would be that whales feeding around Antarctic Peninsula may mate and calve off Ecuador and Colombia, while Magellan Strait humpbacks may breed mainly off Panama and Costa Rica. Additionally, another hypothesis that should not be ruled out is a possible gene flow between whales belonging to Magellan Strait, Panama and Costa Rica, and those from Brazil. However, sampling and genetic analyses need to be conducted in the northernmost breeding areas (Panama and Costa Rica) of the BSG to better test genetically this hypothesis. Another issue is the difficulty to distinguish (except for females with calf) if the whales sampled off Ecuador and/or Colombia are breeding in these areas or just passing in the way to breeding areas located further north. Differences in the occupancy rates have been reported between the two breeding locations off western South America, with a higher concentration of females and calves in the Colombia (Gorgona Island) than in the Ecuador, suggesting that this latter could be a "passing-by" area (Félix and Haase 2001). In order to better understand the whale distribution in the different calving areas, a sampling effort should be performed mainly on arrival and departure of the whales migrating for these areas.

**Funding**

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [to A.L.C.-S.]; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [to S.L.B.]; Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) [to S.L.B.]; Instituto Chico Mendes de Conservação da Biodiversidade/Ministério do Meio Ambiente (ICMBIO/MMA); and Petróleo Brasileiro S.A (PETROBRAS). Special support was given by CAPES for the scholarship award to A.L.C.-S. to conduct part of her doctoral research as a visiting student at Oregon State University (OSU). Biopsy samples were collected off Brazil under permit SISBIO 21489-1.

**Acknowledgements**

This survey was made possible through the generous collaboration of members of the South Pacific Whale Research Consortium (SPWRC). We thank the staff of Instituto Baleia Jubarte and of Fundación Yubarta for collecting Brazilian and Colombian humpback whale samples, respectively, in the field. Also, thanks to INACH, SO-GLOBEC program and IWC for fieldwork in the Antarctic Peninsula. We also thank the colleagues of the Laboratório de Biologia Genômica e Molecular/PUCRS for their help in lab activities, specially to Priscilla Mena Zamberlan and Tiago Ferraz da Silva for laboratory assistance. We are also grateful for laboratory support from Alana Alexander, Renee Gibb, Rebecca Hamner and Beth Slikas during the work at Cetacean Conservation and Genomics Laboratory, Oregon State University. Special thanks to Jennifer Jackson for providing electropherograms of the mtDNA control region sequences for the Brazilian humpback whale samples collected in 1997 and 1998.

**References**

Acevedo-Gutiérrez A, Smultea MA. 1995. First records of humpback whales including calves at Golfo Dulce and Isla del Coco, Costa Rica, suggesting geographical overlap of northern and southern hemisphere populations. Mar. Mammal Sci. 11(4):554–60.

Acevedo J, Acuña P, Olavarría C, Aguayo-Lobo A, Pastene LA. 2004. Report of cetacean surveys in the Magellan Strait in the austral summer 2003/2004. Unpublished paper SC/56/O7 presented to the Scientific Committee of the International Whaling Commission. Sorrento, Italy. 29 June – 10 July.

Acevedo J, Rasmussen K, Félix F, Castro C, Llano M, Secchi ER, Saborio MT, Aguayo-Lobo A, Haase B, Scheidat M, Dalla-Rosa L, Olavarría C, Forestell P, Acuña P, Kaufman G, Pastene L. 2007. Migratory destination of humpback whales from the Magellan Strait feeding ground, Southeast Pacific. Mar Mammal Sci. 23:453-463.

Amos B. 1996. Levels of genetic variability in cetacean populations have probably changed little as a result of human activities. Rep Int Whaling Comm. 46:657-658.

Andriolo A, Martins CCA, Engel MH, Pizzorno JL, Más-Rosa S, Freitas A, Morete ME, Kinas PG. 2006. The first aerial survey to estimate abundance of humpback whale (*Megaptera novaeangliae*) in the breeding ground off Brazil. J Cetacean Res Manage. 8:307-311.

Andriolo A, Kinas PG, Engel MH, Martins CCA. 2010. Humpback whale population estimates and distribution along the Brazilian breeding ground. Endang Spec Resear. 11:233-243.

Baker CS, Herman LM, Perry A, Lawton WS, Straley JM, Wolman AA, Kaufman GD, Winn HE, Hall JD, Reinke JM, Ostman J. 1986. Migratory movement and population structure of humpback whales (*Megaptera novaeangliae*) in the Central and Eastern North Pacific. Mar Ecol Prog Ser. 31:105-119.

Baker CS, Perry A, Bannister JL, Weinrich MT, Abernethy RB, Calambokidis J, Lien J, Lambertsen RH, Urban-Ramirez J, Vasquez O, Clapham PJ, Alling A, O´Brien SJ, Palumbi SR. 1993. Abundant mitochondrial DNA variation and worldwide population structure in humpback whales. Proc Natl Acad Sci USA. 90:8239-8243.

Baker CS, Slade RW, Bannister JL, Abernethy RB, Weinrich MT, Lien J, Urban-Ramirez J, Corkeron P, Calambokidis J, Vasquez O, Palumbi SR. 1994. Hierarchical structure of mitochondrial DNA gene flow among humpback whales, world-wide. Mol Ecol. 3:313-327.

Baker CS, Medrano-Gonzalez, Calambokidis J, Perry A, Pichler F, Rosenbaum H, Straley JM, Urban-Ramirez J, Yamaguchi M, von Ziegesar O. 1998a. Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. Mol Ecol. 7:695-707.

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

Bandelt HJ, Richards M, Macaulay V. 2001. Detecting errors in mtDNA data by phylogenetic analysis. Int J Legal Med. 115:64-69.

Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet. 71:1150-1160.

Bergl RA, Vigilant L. 2007. Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (*Gorilla gorilla diehli*). Mol Ecol. 16:501-516.

Bérubé M, Palsbøll PJ. 1996. Identification of sex in cetaceans by multiplexing using ZFX and ZFY specific primers. Mol Ecol. 5:283-287.

Bérubé M, Jorgensen H, McEwing R, Palsboll PJ. 2000. Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. Mol. Ecol. 9:2181-2183.

Brown MR, Corkeron PJ, Hale PT, Schultz KW, Bryden MM. 1995. Evidence for a sex segregated migration in the humpback whale (*Megaptera novaeangliae*). Proc R Soc Lond B Biol Sci. 259:229-234.

Caballero S, Hamilton H, Jaramillo C, Capella J, Flórez-González L, Olavarría C, Rosenbaum HC, Guhl F, Baker CS. 2001. Genetic characterization of the Colombian pacific coast humpback whale population using RAPD and mitochondrial DNA sequences. Mem Quensl Mus. 47:459-464.

Calambokidis J, Steiger GH, Rasmussen K, Urbán JR, Balcomb KC, Ladrón de Guevara P, Salinas MZ, Jacobsen JK, Baker CS, Herman LM, Cerchio S, Darling JD. 2000. Migratory destinations of humpback whales that feed off California, Oregon and Washington. Mar Ecol Prog Ser. 192:295-304.

Calambokidis J, Steiger GH, Straley JM, Herman LM, Cerchio S, Salden DR, Urbán J, Jacobson JK, von Ziegesar O, Balcomb KC, Gabrielle CM, Dahlheim ME, Uchida S, Ellis G, Miyamura Y, Ladrón de Guevara PP, Yamaguchi M, Sato F, Mizroch SA,

Schlender L, Rasmussen K, Barlow J, Quinn TJ. 2001. Movements and population structure of humpback whales in the North Pacific. Mar Mamm Sci. 17:769-794.

Calambokidis J, Falcone EA, Quinn TJ, Burdin AM, Clapham PJ, Ford JKB, Gabriele CM, LeDuc R, Mattila D, RojasBracho L, Straley JM, Taylor BL, Urbán J, Weller D, Witteveen BH, Yamaguchi M, Bendlin A, Camacho D, Flynn K, Havron A, Huggins J, Malone N, Barlow J, Wade PR. 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).

Capella J, Gibbons J, Flórez-González I, Llano M, Valladares C, Sabaj V, Vilina Y. 2008. Migratory round-trip of individually identified humpback whales at the Strait of Magellan: clues on transit times and phylopatry to destinations. Revista Chilena de Historia Natural. 81:547-60.

Carroll E, Patenaude N, Alexander A, Steel DS, Harcourt R, Childerhouse S, Smith S, Bannister J, Constantine R, Baker CS. 2011. Population structure and individual movement of southern right whales around New Zealand and Australia. Mar Ecol Prog Ser. 432:257–268.

Castro C, Acevedo J, Allen JM, Dalla Rosa L, Flórez-González L, Aguayo-Lobo A, Rassmusen K, Llano M, Garita F, Forestell P, Secchi ER, García-Godos I, Ferina D, Kaufman G, Scheidat M, Pastene L. 2008. Migratory movements of humpback whales (*Megaptera novaeangliae*) between Machalilla National Park, Ecuador and southeast Pacific. Unpublished paper SC/60/SH23 presented to the Scientific Committee of the International Whaling Commission.

Chittleborough RG. 1965. Dynamics of two populations of the humpback whale, *Megaptera novaeangllae* (Borowski). Aust J Mar Freshw Res. 16:33-128.

Clapham PJ, Mead JG. 1999. *Megaptera novaeangliae*. Mammalian Species. 604:1-9.

Clark CW, Clapham PJ. 2004. Acoustic monitoring on a humpabck whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late spring. Proc R Soc Lond B Biol Sci. 271:1051-1057.

Constantine R, Russell K, Gibbs N, Childerhouse S, Baker CS. 2007. Photo-identification of humpback whales (*Megaptera novaeangliae*) in New Zealand waters and their migratory connections to breeding grounds of Oceania. Mar. Mamm Sci. 23(3):715-720.

Craig AS, Herman LM. 1997. Sex differences in site fidelity and migration of humpback whales (*Megaptera novaeangliae*) to the Hawaiian Islands. Can J Zool. 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.

Crawford NG. 2010. SMOGD: software for the measurement of genetic diversity. Mol Ecol Resources. 10:556557.

Cypriano-Souza AL, Fernández GP, Lima-Rosa CAV, Engel MH, Bonatto SL. 2010. Microsatellite genetic characterization of the humpback whale (*Megaptera novaeangliae*) Breeding Ground off Brazil (Breeding Stock A). J Hered. 101(2):189-200.

Dalla Rosa L, Freitas A, Secchi ER, Santos MCO, Engel MH. 2004. An updated comparison of the humpback whale photo-id catalogues from the Antarctic Peninsula and the Abrolhos Bank, Brazil. Unpublished paper SC/56/SH16 presented to the Scientific Committee of the International Whaling Commission.

Dalla Rosa L, Secchi ER, Maia YG, Zerbini AN, Heide-Jorgensen MP. 2008. Movements of satellite-monitored humpback whales on their feeding ground along the Antarctic Peninsula. Polar Biol. 31:771-781.

Darling JD, Cerchio S. 1993. Movement of a humpback whale, *Megaptera novaeangliae*, between Japan and Hawaii. Mar Mamm Sci. 9:84-89.

Darling JD, Souza-Lima RS. 2005. Songs indicate interaction between humpback whale (*Megaptera novaeangliae*) populations in the western and eastern South Atlantic Ocean. Mar Mamm Sci. 21:557-566.

Dawbin W. 1966. The seasonal migratory cycle of the humpback whale. In: Norris KS, editor. Whales, Dolphins and Porpoises. Berkeley: University of California Press; p. 145-170.

Donovan GP. 1991. A review of IWC stock boundaries. In: Hoelzel AR, editor. Genetic ecology of whales and dolphins. Rep Int Whaling Comm Spec Iss. 13:39-68.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resources. 4(2):359-361.

Engel MH, Fagundes NJR, Rosenbaum HC, Leslie MS, Ott PH, Schmitt R, Secchi E, Dalla Rosa L, Bonatto SL. 2008. Mitochondrial DNA diversity of the Southwestern Atlantic humpback whale (*Megaptera novaeangliae*) breeding area off Brazil, and the potential connections to Antartic feeding areas. Conserv Genet. 9:1253-1262.

Engel MH, Martin AR. 2009. Feeding grounds of the western South Atlantic humpback whale population. Mar. Mamm. Sci. 25:964–969.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14(8):2611-2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Lynux and Windows. Mol Ecol Resources. 10:564-567.

Faubet P, Waples RS, Gaggiotti OE. 2007. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. Mol Ecol. 16:1149-1166.

Félix F, Haase B. 2001. The humpback whales off the coast of Ecuador, population parameters and behavior. Revista de Biologia Marina y Oceanografia 36:61-74.

Félix F, Haase B. 2005. Distribution of humpback whales along the coast of Ecuador and management implications. J Cetacean Res Manage. 7(1):21-31.

Félix F, Palacios DM, Caballero S, Haase B, Falconí J. 2006. The 2005 Galápagos Humpback Whale Expedition: a first attempt to assess and characterize the population in the Archipelago. Unpublished paper SC/A06/HW15 presented to the Scientific Committee of the International Whaling Commission.

Félix F, Rasmussen MH, Garita F, Haase B, Simonis A. 2009. Movements of humpback whales between Ecuador and central America, wintering area of the breeding stock G. Unpublished paper SC/61/SH18 presented to the Scientific Committee of the International Whaling Commission.

Félix F, Caballero S, Olavarría C. 2012. Genetic diversity and population structure of humpback whales (*Megaptera novaeanglieae*) from Ecuador based on mitochondrial DNA analyses. J Cetacean Res Manage. 12(1):71-77.

Flórez-González L, Capella J, Haase B, Bravo GA, Félix F, Gerrodette T. 1998. Changes in winter destinations and the northernmost record of southeastern Pacific humpback whales. Mar. Mamm Sci.14(1):189-96.

Freitas AC, Kinas PG, Martins CCA , Engel MH. 2004. Abundance of humpback whales on the Abrolhos Bank wintering ground, Brazil. J Cetacean Res Manage. 3:225-230.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147:915 –925.

Furtado-Neto MA, Monteiro-Neto C, Campos A, Lien J, Carr S. 1998. Are Northern-Hemisphere humpback whales stranding in South Atlantic beaches? Answers from mitochondrial DNA sequences. Abstracts of the 8º Reunião de Trabalhos de Especialistas em Mamíferos Aquáticos da América do Sul and 2º Congresso da Sociedade Latino Americana de Mamíferos Aquáticos, SOLAMAC.

Gambell R. 1973. Sustainable yields: how whales survive. In: Calders N, editor. Nature in the round. London: Weidenfeld and Nicolson. p. 193-202.

Garrigue C, Aguayo-Lobo A, Amante-Helwing VLU. 2002. Movements of humpback whales in Oceania, South Pacific. J Cetacean Res Manage. 4:255-260.

Garrigue C, Dodemont R, Steel D, Baker CS. 2004. Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm low abundance and reproductive autonomy of numpback whales on the wintering grounds of New Caledonia. Mar Ecol Progr Ser. 274:251-262.

Gibbons J, Capella JJ, Valladares C. 2003. Rediscovery of a humpback whale (*Megaptera novaeangliae*) feeding ground in the Straits of Magellan, Chile. J Cetacean Res Manage. 5:203-208.

Goudet J, 2002. FSTAT: a program to estimate and test gene diversities and fixation indices. Lausanne: Institute of Ecology.

Greenwood PJ. 1983. Mating systems and the evolutionary consequences of dispersal. In: Swingland IR, Greenwood PJ. editors. The Ecology of Animal Movement. Oxford: Clarendon Press. p. 116-131.

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

Hammer RM, Pichler FB, Heimeier D, Constantine R, Baker CS. 2012. Genetic differentiation and limited gene flow among fragmented populations of New Zealand endemic Hector's and Maui's dolphins. Conserv Genet. 13: 987-1002.

Heller R, Okello JBA, Siegismund H. 2010. Can small wildlife conservancies maintain genetically stable populations of large mammals? Evidence for increased genetic drift in geographically restricted populations of Cape buffalo in East Africa. Mol Ecol. 19:1324-1334.

Hoelzel AR. 1998. Genetic structure of cetacean populations in sympatry, parapatry, and mixed assemblages: Implications for conservation policy. J Hered. 89:451-458.

Hollatz C, Vilaça ST, Redondo RAF, Marmontel M, Baker CS, Santos FR. 2011. The Amazon River system as an ecological barrier driving genetic differentiation of the pink dolphin (*Inia geoffrensis*). Biological Journal of the Linnean Society. 102:812–827.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. Mol Ecol Resources. 9:1322–1332.

International Whaling Commission, 2005. Report of the Scientific Committee. Annex H. Report of the sub-committee on other Southern Hemisphere whale stocks. J Cetacean Res Manage. 7:235-46.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801-1806.

Jost L. 2008. $G_{ST}$ and its relatives do not measure differentiation. Mol Ecol. 17:4015-4026.

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol. 16:1099-1106.

Lambeck K, Esat TM, Potter E-K. 2002. Links between climate and sea levels for the past three million years. Nature. 419:199-206.

Lambertsen RH. 1987. A biopsy system for large whales and its use for cytogenetics. J Mamm. 68:443-445.

Lecis R, Pierpaoli M, Biro ZS, Szemethy L, Ragni B, Vercillo F, Randi E. 2006. Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. Mol Ecol. 15:119-131.

Lodi L. 1994. Ocorrências de baleias-jubarte, *Megaptera novaeangliae*, no Arquipélago de Fernando de Noronha, incluindo um resumo de registros de capturas no Nordeste do Brasil. Biotemas. 7:116-123.

Mackintosh NA 1965. The stocks of whales. Vol. Fishing News (Books) Ltd., London.

Magalhães FA, Tosi CH, Garri RG, Chellappa S, Silva FL. 2008. Cetacean diversity on the Parnaiba Delta, Maranhão state, northeastern Brazil. Braz J Biol. 68(3):545-551.

Martins CCA, Morete ME, Engel MH, Freitas A, Secchi ER, Kinas PG. 2001. Aspects of habitat use patterns of humpback whales in the Abrolhos bank, Brazil, breeding ground. Mem Queensl Mus. 47:563-570.

Medrano-González L, Baker CS, Robles-Saavedra MR, Murrell J, Vazquez-Cuevas MJ, Congdon BC, Stranley JM, Calambokidis J, Urban-Ramirez J, Flórez-González L, Olavarría C, Aguayo-Lobo A, Nolasco-Soto J, Juarez-Salas RA, Villavicencio-Llamosas K. 2001. Trans-oceanic population genetics structure of humpback whales in the North and South Pacific. Mem Queensl Mus. 47:465-479.

Meirelles ANO, Monteiro-Neto C, Martins AMA, Costa AF, Barros HMDR, Alves MDO. 2009. Cetacean strandings on the coast of Ceará, northeastern Brazil (1992-2005). J Mar Biol Assoc United Kingdom. 89(5):1083-1090.

Mikhalev YA. 1997. Humpback whales *Megaptera novaeangliae* in the Arabian Sea. Mar Ecol Prog Ser. 149:13-21.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89:583-590.

Noad MJ, Cato, DH, Bryden MM, Jenner MN, Jenner KCS. 2000. Cultural revolution in whale songs. Nature. 408:537-537.

Olavarría C, Baker CS, Medrano L, Aguayo A, Caballero S, Flórez-González L, Capella J, Rosenbaum HC, Garrigue C, Greaves J, Jenner M, Jenner C, Bannister JL. 2000. Stock identity of Antarctic Peninsula Humpback whales inferred from mtDNA variation. Unpublished paper SC/52/IA15 to the Scientific Committee of the International Whaling Commission.

Olavarría, C., Aguayo, A., Acevedo, J., Medrano, L., Thiele, D. and Baker, C.S. 2006. Genetic differentiation between two feeding areas of the Eastern South Pacific humpback whale population: update on SC/57/SH3. Unpublished paper SC/A06/HW29 presented to the Scientific Committee of the International Whaling Commission.

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 humpback whales throughout the South Pacific and the origin of the eastern Polynesian breeding grounds. Mar Ecol Prog Ser. 330:257-268.

Pacheco AS, Silva S, Alcorta B. 2009. Winter distribution and group composition of humpback whales (*Megaptera novaeangliae*) off Northern Peru. Lat Am J Aquat Mamm. 7(1-2):33-38.

Paetkau D, Slade R, Burden M, Estoup A. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol. 13:55-65.

Palsbøll PJ, Vader A, Bakke I, El-Gewely MR. 1992. Determination of gender in cetaceans by polymerase chain reaction. Can J Zool. 70:2166-2170.

Palsbøll PJ, Clapham PJ, Mattila DK, Larsen F, Sears R, Siegismund HR, Sigurjónsson J, Vásquez O, Artander P. 1995. Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. Mar Ecol Prog Ser. 116:1-10.

Palsbøll PJ, Allen J, Bérubé M, Clapham PJ, Feddersen TP, Hammond PS, Hudson RR, Jørgensen H, Katona S, Larsen AH, Larsen F, Lien J, Mattila DK, Sigurjonsson J, Sears

R, Smith T, Sponer R, Stevick P, Øien N. 1997a. Genetic Tagging of Humpback Whales. Nature. 388:767-769.

Palsbøll PJ, Bérubé M, Jorgensen H. 1997b. Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. Mol Ecol. 6:893-895.

Palsbøll PJ, Peery MZ, Bérubé M. 2010. Detecting populations in the 'ambiguous' zone: kinship-based estimation of population structure at low genetic divergence. Mol Ecol Resources. 10:797-805.

Palumbi SR, Baker CS. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Bio Evol. 11:426-435.

Piry S, Alapetite A, Cornuet JM, Paektau D, Baudouin L, Estoup A. 2004. GeneClass2: a software for genetics assignment and first generation migrants detection. J Hered. 95:536-539.

Pomilla C, Rosenbaum HC. 2005. Against the current: an inter-oceanic whale migration event. Biol Lett. 1:476-479.

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

Pretto DJ, Andrade MCM, Oliveira JM, Oliveira MGA. 2009. First record of a humpback whale, *Megaptera novaeangliae* (Borowski, 1781), stranding in Pará State, Northern coast of Brazil. Braz J Biol. 69(4):1207-1208.

Pritchard JK, Stephens M, Donnely P. 2000. Inference of a population structure using multilocus genotype data. Genetics. 155:945-959.

Pritchard JK, Wen X, Falush D. 2007. Documentation for structure software: version 2.2. Available from http://pritch.bsd.uchicago.edu/software

Ramírez P. 1988. La ballena jorobada *Megaptera novaeangliae* en la costa norte del Perú. Periodos 1961-1966 y 1975-1985. Bol. Lima. 56:91-96.

Rannala B, Mountain JL. 1997. Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA. 94:9197-9221.

Rasmussen K, Palacios D, Calambokidis J, Saborio MT, Dalla Rosa L, Secchi ER, Steiger GH, Allen JM, Stone G. 2007. Southern Hemisphere humpback whales wintering off Central America: insights from water temperature into the longest mammalian migration. Biology Letters. 3(3):302-05.

Rice WR. 1989. Analyzing table of statistical tests. Evolution. 43:223-225.

Robbins J, Dalla Rosa L, Allen JM, Mattila D, Secchi ER, Friedlaender AS, Stevick PT, Nowacek DP, Steel D. 2011. Return movement of a humpback whale between the Antarctic Peninsula and American Samoa: a seasonal migration record. Endang Species Res. 13:117-121.

Roman J, Palumbi S. 2003. Whales before whaling in North Atlantic. Science. 301:508-510.

Rosenbaum HC, Pomilla C, Mendez M, Leslie MS, Best PB, Findlay KP, Minton G, Ersts PJ, Collins T, Engel MH, Bonatto SL, Kotze DPGH, Meÿer M, Barendse J, Thornton M, Razafindrakoto Y, Ngouessono S, Vely M, Kiszka J. 2009. Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. PLoS ONE. 4:e7318.

Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes. 4(1):137-138.

Rossi-Santos MR, Neto ES, Baracho CG, Cipolotti SR, Marcovaldi E, Engel MH. 2008. Occurrence and distribution of humpback whales (Megaptera novaeangliae) on the North coast of the State of Bahia, Brazil, 2000-2006. ICES J Mar Sci. 65:1-7.

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

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. 2012. Low levels of genetic differentiation characterize Australian humpback whale (*Megaptera novaeangliae*) populations. Unpublished paper SC/64/SH15 to the Scientific Committee of the International Whaling Commission.

Siciliano S. 1997. Características da população de baleias-jubarte (*Megaptera novaeangliae*) da costa brasileira, com especial referência aos Bancos de Abrolhos. MS Thesis, Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro.

Siciliano S, Moura JF, Filgueiras HR, Rodrigues PP, Leite Jr. NO. 2012. Sightings of humpback whales on the Vitória-Trindade Chain and around Trindade Island, Brazil. Brazilian Journal of Oceanography. 60(3): 455-459.

Slatkin M. 1991. Inbreeding coefficients and coalescence times. Genetical Research. 58:167-175.

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

Slijper EJ. 1962. Whales. 1st English Edn. Hutchinson & Co., London. 475pp.

Slijper EJ. 1965. A hundred years of modern whaling. Netherlands Commmission for International Nature Protection, Amsterdam. 43pp.

Sremba AL, Martin AR, Baker CS. 2010. Genetic approach to species identification of whale bones from South Georgia Island Whaling Stations. Unpublished paper SC/62/SH19 to the Scientific Committee of the International Whaling Commission.

Sremba AL, Hancock-Hanser B, Branch TA, LeDuc RL, Baker CS. 2012. Circumpolar diversity and geographic differentiation of mtDNA in the critically endangered Antarctic blue whale (*Balaenoptera musculus intermedia*). PLoS ONE. 7:e32579.

Steel D, Garrigue C, Caballero S, Thiele D, Poole M, Paton D, Hauser N, Clapham P, Olavarría C, Donoghue M, Flórez-González, Baker CS, Constantine R. 2008. Migratory connections between humpback whales from South Pacific breeding grounds and Antarctic feeding areas based on genotype matching. Unpublished paper SC/60/SH13 presented to the Scientific Committee of the International Whaling Commission.

Stevick PT, Allen J, Bérubé MC. 2003. Segregation of migration by feeding ground origin in North Atlantic humpback whales (*Megaptera novaeangliae*). J Zool. 259:231-237.

Stevick PT, Aguayo A, Allen J, Avila IC, Capella J, Castro C, Chater K, Engel M, Félix F, Flórez-González L, Freitas A, Hasse B, Llano M, Lodi L, Muñoz E, Olavarría C, Secchi E, Scheidat M, Siciliano S. 2004. A note on the migrations of individually identified humpback whales between the Antarctic Peninsula and South America. J Cetacean Res Manage. 6:109–113.

Stevick PT, Allen J, Clapham PJ, Katona SK, Larsen F, Lien J, Mattila DK, Sears PJPR, Sigurjónsson J, Smith TD, Vikingsson G, Øien N, Hammond PS. 2006a. Population spatial structuring on the feeding grounds in North Atlantic humpback whales (*Megaptera novaeangliae*). J Zool. 270:244-255.

Stevick PT, Pacheco de Godoy L, McOsker M, Engel MH, Allen A. 2006b. A note on the movement of a humpback whale from Abrolhos Bank, Brazil to South Georgia. J Cetacean Res Manage. 8:297–300.

Stevick PT, Neves MC, Johansen F, Engel MH, Allen J, Marcondes MCC, Carlson C. 2010. A quarter of a world away: female humpback whale moves 10,000 km between breeding areas. Biology Letters. 7:299-302.

Stevick PT, Allen JM, Engel MH, Félix F, Haase B, Neves M. 2011. First record of inter-oceanic movement of a humpback whale between Atlantic and Pacific breeding grounds off South America. Unpublished paper SC/63/SH4 to the Scientific Committee of the International Whaling Commission

Stone GS, Flórez-González L, Katona S. 1990. Whale migration record. Nature. 346: 705-706.

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

Tønnessen JN, Johnsen AO. 1982. The history of modern whaling. Berkeley: University of California Press.

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

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

Valsecchi E, Corkeron PJ, Galli P, Sherwin W, Bertorelle G. 2010. Genetic evidence for sex-specific migratory behaviour in western South Pacific humpback whales. Mar. Ecol. Prog. Ser. 398:275-86.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping erros in microsatelite data. Mol Ecol Notes. 4:535-538.

Waldick RC, Brown MW, White BN. 1999. Characterization and isolation of microsatellite loci from the endangered North Atlantic right whale. Mol Ecol. 8:1763-1765.

Waples RS, Gaggiotti O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol. Ecol. 15:1419-1439.

Wedekin LL, Rossi-Santos M, Baracho C, Cypriano-Souza AL, Simões-Lopes PC. in press. Cetacean records along a coastal-offshore gradient in the Vitória-Trindade Chain, western South Atlantic Ocean. Brazilian Journal of Biology.

Weinrich MT, Rosenbaum HC, Baker CS, Blackmer AL, Whitehead H. 2006. The influence of maternal lineages on social affiliations among humpback whales (*Megaptera novaeangliae*) on their feeding grounds in the southern Gulf of Maine. J Hered. 97:226-234.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the Analysis of Population Structure. Evolution. 38:1358-1370.

Wilberg MJ, Dreher BP. 2004. GENECAP: a program for analysis of multilocus genotype data for non-invasive sampling and capture-recapture population estimation. Mol Ecol Notes. 4:783-785.

Zerbini AN, Andriolo A, da Rocha JM, Simões-Lopes PC, Siciliano S, Pizzorno JL, Waite JM, DeMaster DP, VanBlaricon GR. 2004. Winter distribution and abundance of humpback whales (*Megaptera novaeangliae*) off Northeastern Brazil. J. Cetacean Res. Manage. 6:101-107.

Zerbini AN, Andriolo A, Heide-Jorgensen MP, Pizzorno JL, Maia YG, VanBlaricom GR, DeMaster DP, Simões-Lopes PC, Moreira S, Bethlem C. 2006. Satellite-monitored movements of humpback whales *Megaptera novaeangliae* in the Southwest Atlantic Ocean. Mar Ecol Prog Ser. 313:295-304.

Zerbini AN, Andriolo A, Danilewicz D, Heide-Jorgensen MP, Gales N, Clapham PJ. 2011. An update on research on migratory routes and feeding destinations of Southwest Atlantic humpback whales. Unpublished paper SC/63/SH23 presented to the Scientific Committee of the International Whaling Commission.

**Figure Legends**

Figure 1. Map of the regions studied: the two breeding grounds off eastern and western coast of South America, Brazil (Breeding stock A) and Colombia (Breeding stock G), respectively, and the feeding area around Antarctic Peninsula.

Figure 2. STRUCTURE barplot of the proportional membership (q) of each individual of humpback whale in each cluster for $K = 2$ (a) and $K = 3$ (b) without use of location prior, and for $K = 2$ (c) and $K = 3$ (d) with sampling location prior. Each individual is represented by a vertical bar broken into colored segments with the length indicating the coefficient of membership to each population. BSA - Breeding stock A (blue or yellow), BSG - Breeding stock G and AP - Antarctic Peninsula (orange). Black line represents the boundary between individuals sampled in each of the three regions.

Figure 3. Assignment of individuals to the Breeding stock A (red) or Breeding stock G (green) using STRUCTURE with the prior population information (assuming $K = 2$, BSA and BSG - Colombia plus Antarctic Peninsula) incorporated. Arrows indicate individuals with admixed ancestry ($0.2 < q > 0.8$) and potential migrants ($q > 0.9$).

Figure S1. *Megaptera novaeangliae*. Positions of the 67 variable sites defining 77 mtDNA control region haplotypes (465 bp) detected in 370 humpback whales sampled in two breeding ground (BSA, $n = 158$; BSG, $n = 130$) and one feeding ground (Antarctic Peninsula, $n = 82$). Position 1 of alignment corresponds to position 22 in Engel et al. (2008) and Rosenbaum et al. (2009). Dots (.) indicate matches with reference sequence (HBA002 or SP1) and dashes (-) indicate insertion/deletion. The frequencies of haplotypes are shown for each of the 3 sampling regions. Haplotype nomenclature follows Olavarria et al. (2006, 2007), Engel et al. (2008) and Rosenbaum et al. (2009).

Figure S2. Mean log-likelihood L($K$) and $\Delta K$ calculated using STRUCTURE HARVESTER for one to five clusters ($K$) based on ten independent STRUCTURE runs, with a maximum value achieved at $K = 2$.

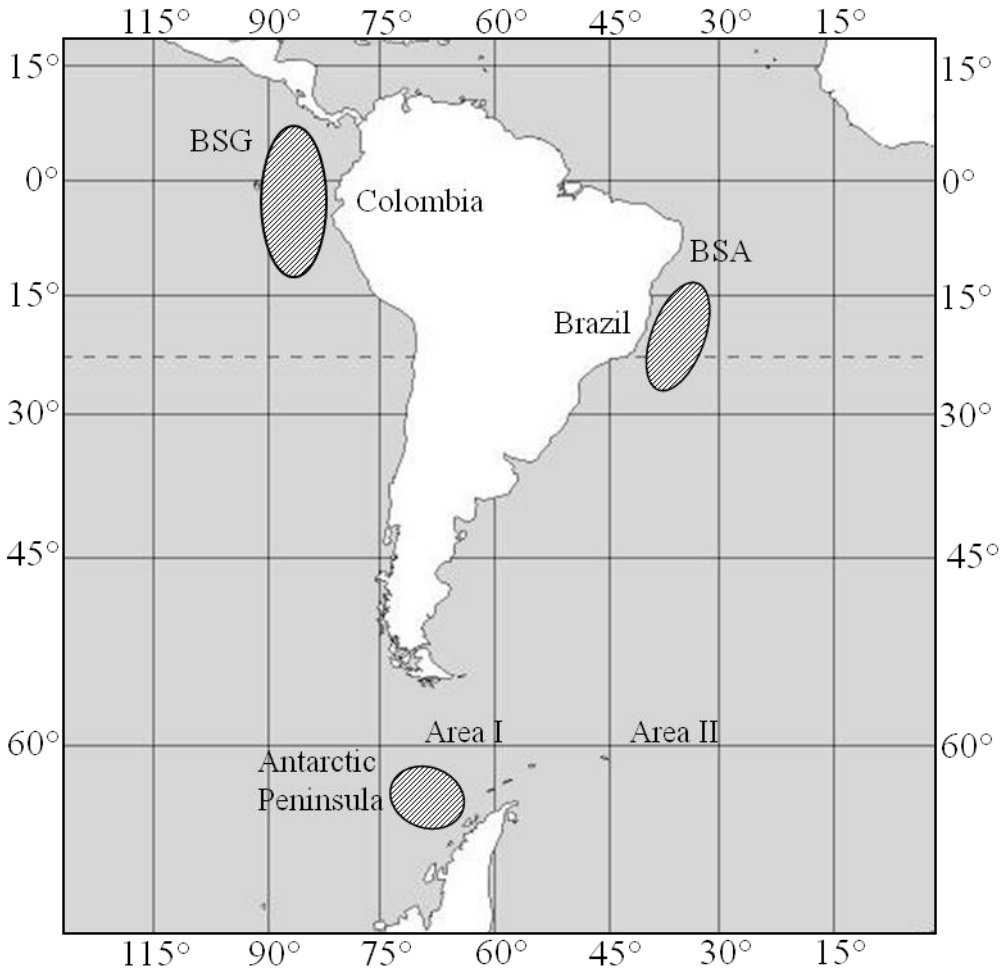

115°
90°
75°
60°
45°
30°
15°
15°
0°
15°
30°
45°
60°
BSG
Colombia
BSA
Brazil
Area I
Area II
Antarctic
Peninsula

Figure 1

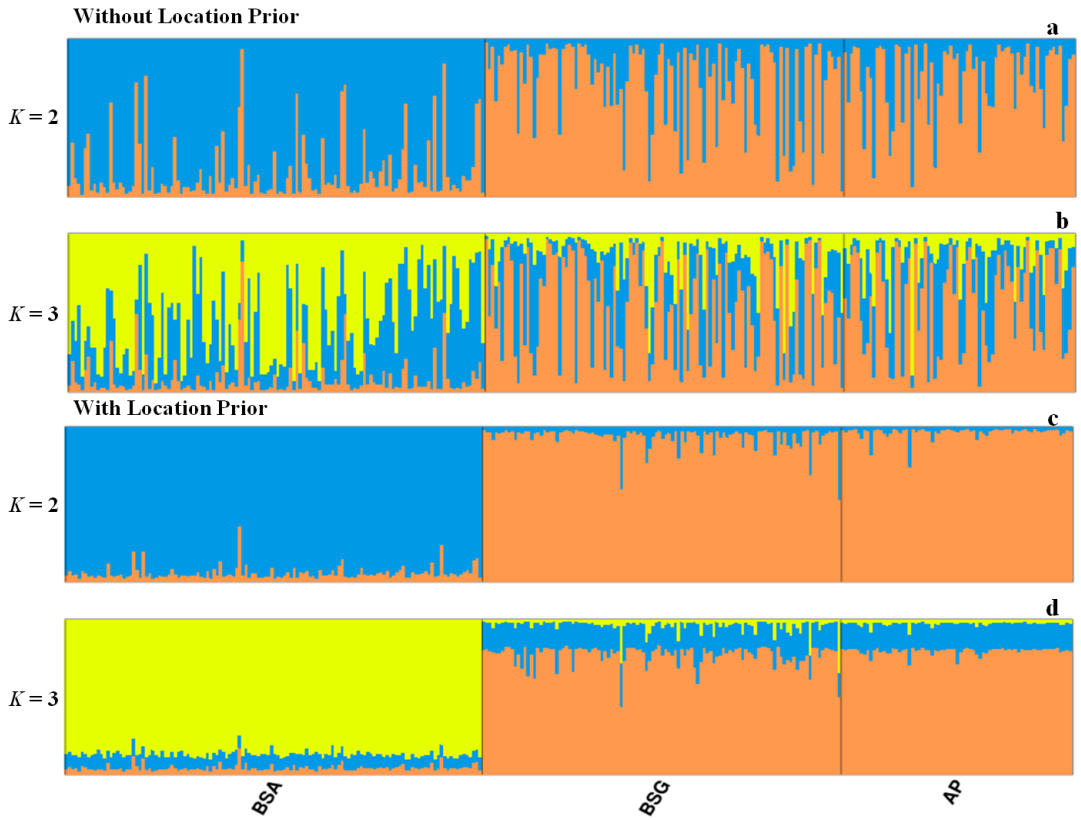



Figure 2

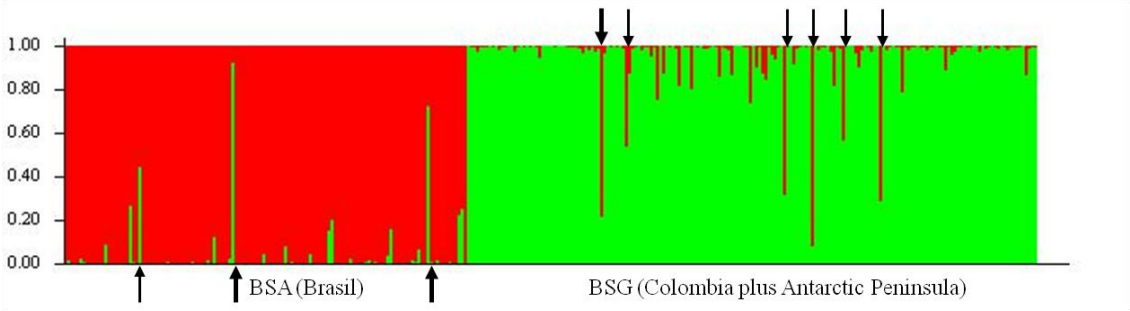
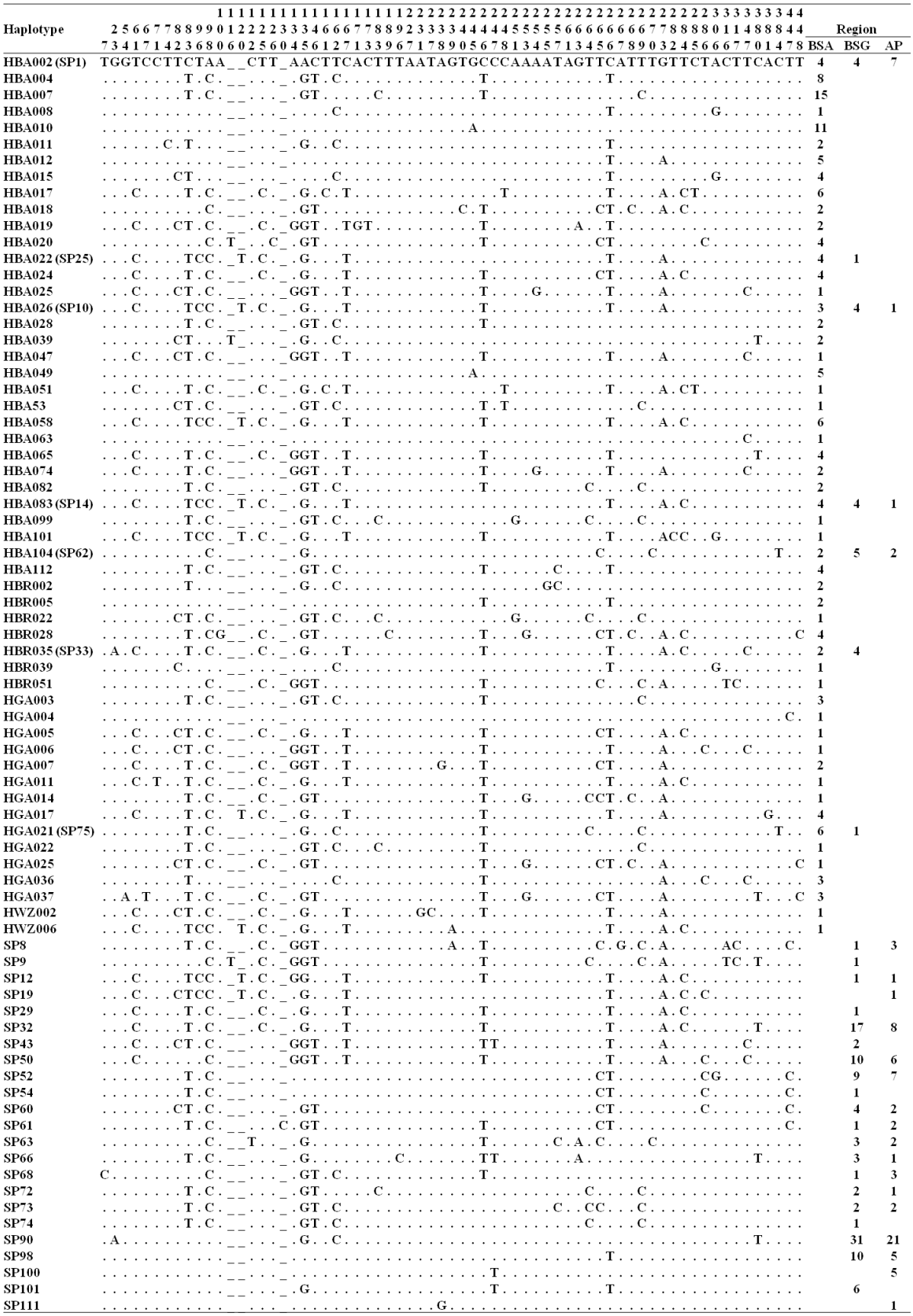
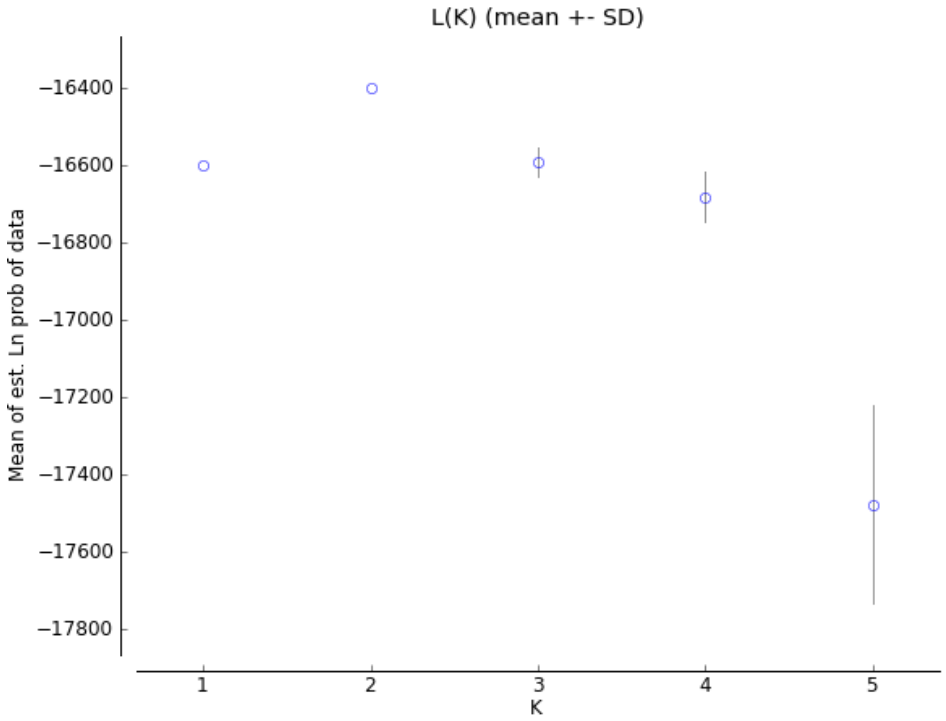



Figure 3



| Haplotype | BSA | Region BSG | AP |
|---|---|---|---|
| HBA002 (SP1) | 4 | 4 | 7 |
| HBA004 | 8 | | |
| HBA007 | 15 | | |
| HBA008 | 1 | | |
| HBA010 | 11 | | |
| HBA011 | 2 | | |
| HBA012 | 5 | | |
| HBA015 | 4 | | |
| HBA017 | 6 | | |
| HBA018 | 2 | | |
| HBA019 | 2 | | |
| HBA020 | 4 | | |
| HBA022 (SP25) | 4 | 1 | |
| HBA024 | 4 | | |
| HBA025 | 1 | | |
| HBA026 (SP10) | 3 | 4 | 1 |
| HBA028 | 2 | | |
| HBA039 | 2 | | |
| HBA047 | 1 | | |
| HBA049 | 5 | | |
| HBA051 | 1 | | |
| HBA53 | 1 | | |
| HBA058 | 6 | | |
| HBA063 | 1 | | |
| HBA065 | 4 | | |
| HBA074 | 2 | | |
| HBA082 | 2 | | |
| HBA083 (SP14) | 4 | 4 | 1 |
| HBA099 | 1 | | |
| HBA101 | 1 | | |
| HBA104 (SP62) | 2 | 5 | 2 |
| HBA112 | 4 | | |
| HBR002 | 2 | | |
| HBR005 | 2 | | |
| HBR022 | 1 | | |
| HBR028 | 4 | | |
| HBR035 (SP33) | 2 | 4 | |
| HBR039 | 1 | | |
| HBR051 | 1 | | |
| HGA003 | 3 | | |
| HGA004 | 1 | | |
| HGA005 | 1 | | |
| HGA006 | 1 | | |
| HGA007 | 2 | | |
| HGA011 | 1 | | |
| HGA014 | 1 | | |
| HGA017 | 4 | | |
| HGA021 (SP75) | 6 | 1 | |
| HGA022 | 1 | | |
| HGA025 | 1 | | |
| HGA036 | 3 | | |
| HGA037 | 3 | | |
| HWZ002 | 1 | | |
| HWZ006 | 1 | | |
| SP8 | | 1 | 3 |
| SP9 | | 1 | |
| SP12 | | 1 | 1 |
| SP19 | | | 1 |
| SP29 | | 1 | |
| SP32 | | 17 | 8 |
| SP43 | | 2 | |
| SP50 | | 10 | 6 |
| SP52 | | 9 | 7 |
| SP54 | | 1 | |
| SP60 | | 4 | 2 |
| SP61 | | 1 | 2 |
| SP63 | | 3 | 2 |
| SP66 | | 3 | 1 |
| SP68 | | 1 | 3 |
| SP72 | | 2 | 1 |
| SP73 | | 2 | 2 |
| SP74 | | 1 | |
| SP90 | | 31 | 21 |
| SP98 | | 10 | 5 |
| SP100 | | | 5 |
| SP101 | | 6 | |
| SP111 | | | 1 |

Figure S1


L(K) (mean +- SD)
Mean of est. Ln prob of data
−16400
−16600
−16800
−17000
−17200
−17400
−17600
−17800
1
2
3
4
5
K

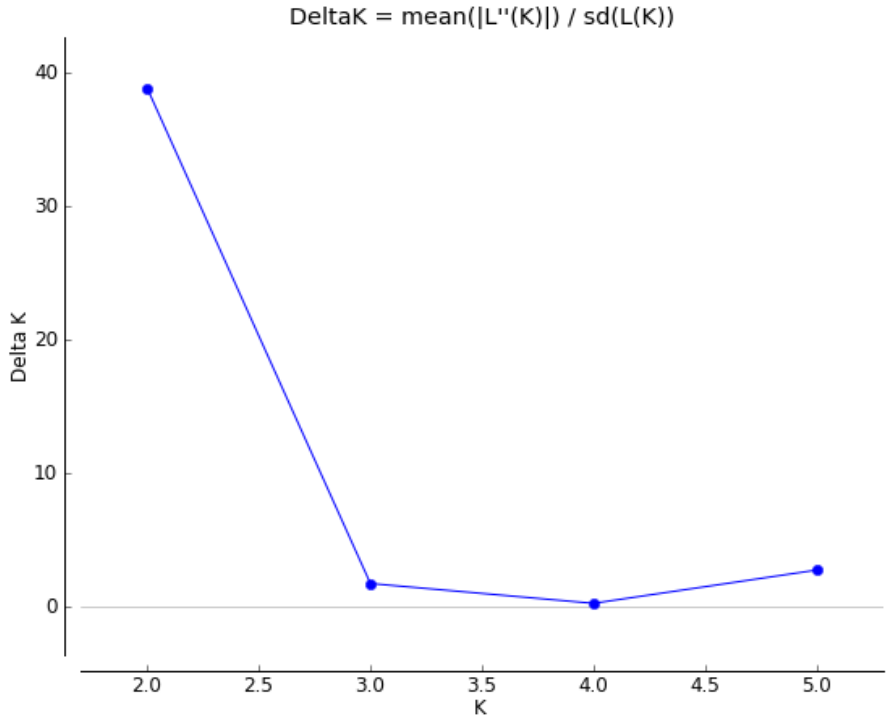

DeltaK = mean(|L''(K)|) / sd(L(K))
Delta K
40
30
20
10
0
2.0
2.5
3.0
3.5
4.0
4.5
5.0
K

Figure S2

Table 1. *Megaptera novaeangliae*. Summary of sampling years, sample size and mtDNA diversity of the humpback whales sampled in two breeding grounds (Brazil and Colombia) and one feeding area (Antarctic Peninsula). Includes number of haplotypes, number of variable sites, haplotype (*h*) and nucleotide ($\pi$) diversities. For *h* and $\pi$, standard deviations are shown in parentheses. Numbers of males (M) and females (F) with their respective mtDNA variability are included in parentheses. Duplicate samples were removed from the analysis.

| Stock | Region | Sampling years | All samples (M/F) | Number of haplotypes | Variable sites | *h* (SD) | *Π* % (SD) |
|---|---|---|---|---|---|---|---|
| BSA | Southwestern Atlantic/ Brazil | 1997-2001 | 158 (83/69) | 54 (43/32) | 60 (56/53) | 0.973 (0.004) (0.974 (0.007)/0.966 (0.008)) | 2.00 (0.047) (1.90 (0.061)/1.90 (0.078)) |
| BSG | Southeastern Pacific/ Colombia | 1991-1999 | 130 (79/36) | 27 (25/11) | 41 (40/33) | 0.906 (0.015) (0.928 (0.014)/0.783 (0.063)) | 1.80 (0.052) (1.90 (0.061)/1.70 (0.161)) |
| BSG | Antarctic Peninsula | 1990, 1994, 1996-1999 | 82 (38/44) | 21 (16/16) | 40 (39/35) | 0.902 (0.020) (0.895 (0.036)/0.915 (0.023)) | 1.80 (0,095) (2.00 (0.149)/1.60 (0.001)) |
| | Overall | 1990-2011 | 370 (200/149) | 77 (66/48) | 66 (64/59) | 0.962 (0.005) (0.967 (0.005)/0.953 (0.010)) | 1.90 (0.032) (2.00 (0.042)/1.90 (0.050)) |

Table 2. Summary of microsatellite diversity (16 loci) for humpback whales sampled in the two breeding grounds (Brazil and Colombia) and one feeding area (Antarctic Peninsula). Includes number of individuals (*n*), number of alleles (*K*), observed ($H_O$) and expected ($H_E$) heterozygosity, and inbreeding coefficient ($F_{IS}$) for each locus. Probability of identity for each population was included. Duplicate samples were removed from the analysis.

| | BSA (Brazil) | | | | | BSG (Colombia) | | | | | Antarctic Peninsula | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Locus | *N* | *K* | $H_O$ | $H_E$ | $F_{IS}$ | *n* | *K* | $H_O$ | $H_E$ | $F_{IS}$ | *n* | *K* | $H_O$ | $H_E$ | $F_{IS}$ |
| EV1 | 129 | 4 | 0.550 | 0.537 | -0.026 | 111 | 4 | 0.595 | 0.586 | -0.014 | 72 | 4 | 0.458 | 0.527 | 0.130 |
| EV14 | 115 | 9 | 0.739 | 0.712 | -0.038 | 112 | 9 | 0.821 | 0.796 | -0.032 | 72 | 9 | 0.736 | 0.793 | 0.072 |
| EV21 | 128 | 5 | 0.742 | 0.725 | -0.024 | 109 | 5 | 0.679 | 0.685 | 0.009 | 70 | 5 | 0.743 | 0.684 | -0.087 |
| EV37 | 129 | 17 | 0.853 | 0.925 | 0.078 | 112 | 15 | 0.857 | 0.906 | 0.054 | 71 | 15 | 0.944 | 0.899 | -0.051 |
| EV94 | 128 | 10 | 0.820 | 0.816 | -0.005 | 111 | 9 | 0.712 | 0.823 | 0.136 | 70 | 9 | 0.771 | 0.804 | 0.040 |
| EV96 | 130 | 16 | 0.915 | 0.860 | -0.065 | 111 | 13 | 0.856 | 0.866 | 0.012 | 73 | 11 | 0.822 | 0.855 | 0.039 |
| EV104 | 129 | 4 | 0.372 | 0.406 | -0.084 | 111 | 4 | 0.495 | 0.495 | -0.001 | 70 | 5 | 0.514 | 0.472 | -0.089 |
| GT23 | 121 | 9 | 0.843 | 0.827 | -0.019 | 108 | 8 | 0.778 | 0.760 | -0.024 | 71 | 7 | 0.775 | 0.754 | -0.028 |
| GT211 | 127 | 10 | 0.780 | 0.811 | 0.039 | 110 | 11 | 0.791 | 0.825 | 0.041 | 66 | 9 | 0.833 | 0.835 | 0.003 |
| GT575 | 124 | 16 | 0.879 | 0.872 | -0.008 | 112 | 13 | 0.795 | 0.802 | 0.009 | 72 | 12 | 0.736 | 0.772 | 0.047 |
| rw4-10 | 126 | 12 | 0.817 | 0.848 | 0.036 | 111 | 12 | 0.793 | 0.849 | 0.067 | 47 | 10 | 0.766 | 0.838 | 0.087 |
| rw31 | 128 | 9 | 0.695 | 0.667 | -0.043 | 110 | 8 | 0.545 | 0.631 | 0.136 | 66 | 8 | 0.561 | 0.613 | 0.086 |
| rw48 | 129 | 6 | 0.736 | 0.730 | -0.009 | 112 | 5 | 0.661 | 0.719 | 0.082 | 71 | 5 | 0.775 | 0.754 | -0.027 |
| GATA28 | 130 | 13 | 0.638 | 0.588 | -0.086 | 110 | 9 | 0.400 | 0.416 | 0.040 | 57 | 8 | 0.316 | 0.315 | -0.001 |
| GATA417 | 130 | 17 | 0.962 | 0.920 | -0.045 | 91 | 13 | 0.923 | 0.906 | -0.019 | 47 | 13 | 0.957 | 0.884 | -0.084 |
| 464/465 | 130 | 6 | 0.585 | 0.600 | 0.026 | 106 | 5 | 0.491 | 0.605 | 0.190 | 71 | 5 | 0.577 | 0.597 | 0.033 |
| Average | | 10.19 | 0.745 | 0.740 | -0.007 | | 8.94 | 0.699 | 0.729 | 0.041 | | 8.44 | 0.705 | 0.712 | 0.010 |
| P(ID) | | $1.30 \times 10^{-18}$ | | | | | $1.02 \times 10^{-17}$ | | | | | $8.17 \times 10^{-17}$ | | | |

Table 3. AMOVA results among humpback whales from Brazil (BSA), Colombia (BSG) and Antarctic Peninsula using mtDNA control region sequences (464 bp) and 16 microsatellite loci. For mtDNA data were estimated molecular distances ($\Phi_{ST}$) and haplotype frequencies ($F_{ST}$). For microsatellites data were estimated $F_{ST}$, $R_{ST}$ and $D_{EST}$ measures of differentiation.

| | Mitocondrial DNA | | Microsatellites | | |
|---|---|---|---|---|---|
| | Overall $F_{ST}$ ($P$) | Overall $\Phi_{ST}$ ($P$) | Overall $F_{ST}$ ($P$) | Overall $R_{ST}$ ($P$) | Overall $D_{EST}$ |
| All samples | | | | | |
| BSA x BSG x Antarctic Peninsula | **0.044** (0.000) | **0.04**1 (0.000) | **0.008** (0.000) | 0.005 (0.079) | 0.018 |
| BSA x (BSG + Antarctic Peninsula) | **0.058** (0.000) | **0.053** (0.000) | **0.011** (0.000) | **0.008** (0.003) | 0.024 |
| Males | | | | | |
| BSA x BSG x Antarctic Peninsula | **0.041** (0.000) | **0.028** (0.001) | 0.006 (0.238) | -0.000 (0.944) | 0.011 |
| BSA x (BSG + Antarctic Peninsula) | **0.052** (0.000) | **0.033** (0.000) | **0.010** (0.000) | 0.006 (0.164) | 0.025 |
| Females | | | | | |
| BSA x BSG x Antarctic Peninsula | **0.072** (0.000) | **0.070** (0.000) | 0.008 (0.206) | 0.003 (0.646) | 0.010 |
| BSA x (BSG + Antarctic Peninsula) | **0.087** (0.000) | **0.085** (0.000) | **0.010** (0.002) | 0.007 (0.204) | 0.020 |

Overall values are shown for all samples, males and females. Significant P values (<0.05) are highlighted in bold.

Table 4. Differentiation pairwise test among two humpback whale breeding grounds (Brazil and Colombia) and one feeding area (Antarctic Peninsula) based on $F_{ST}$ (below diagonal) and $\Phi_{ST}$ (above diagonal) values for mtDNA control region haplotypes (top), and $F_{ST}$ (below diagonal) and $R_{ST}$ (above diagonal) measures for microsatellite loci (bottom).

| Region (stock) | Brazil (BSA) | Colombia (BSG) | Antarctic Peninsula |
|---|---|---|---|
| **mtDNA** | | | |
| N | 158 | 130 | 82 |
| Brazil (BSA) | - | **0.0460** | **0.0627** |
| Colombia (BSG) | **0.0572** | - | 0.0045 |
| Antarctic Peninsula | **0.0589** | 0.0012 | - |
| **Microsatellites** | | | |
| N | 130 | 112 | 72 |
| Brazil (BSA) | - | **0.0096** | 0.0047 |
| Colombia (BSG) | **0.0115** | - | 0.0007 |
| Antarctic Peninsula | **0.0111** | 0.0006 | - |

Significant values are highlighted in bold ($P < 0.001$).

Table 5. Sex-biased dispersal test results of humpback whales from breeding stock A (Brazil) and breeding stock G (Colombia plus Antarctic Peninsula).

| | *mAIc* | *vAIc* | $F_{IS}$ |
|---|---|---|---|
| Male | -0.179 | 14.54 | -0.010 |
| Female | 0.229 | 12.65 | 0.009 |
| P value | 0.354 | 0.426 | 0.251 |

Differences in mean corrected assignment index (*mAIc*), variance of corrected assignment index (*vAIc*), and inbreeding coefficient ($F_{IS}$) were tested for significance using 10 000 permutations.

Table 6. Sex-specific pairwise differentiation among humpback whales sampled in Brazil, Colombia and Antarctic Peninsula based on $F_{ST}$ and $\Phi_{ST}$ (in parentheses) values for mtDNA control region haplotypes (top), and $F_{ST}$ value for microsatellite loci (bottom).

| Region (stock) | | Brazil (BSA) | Colombia (BSG) | Antarctic Peninsula |
|---|---|---|---|---|
| **mtDNA** | females | | | |
| Males | | 83 | 79 | 38 |
| Brazil (BSA) | 69 | - | **0.119 (0.101)** | **0.056 (0.065)** |
| Colombia (BSG) | 36 | **0.046 (0.025)** | - | 0.020 (0.017) |
| Antarctic Peninsula | 44 | **0.060 (0.050)** | 0.006 (0.010) | - |
| **Microsatellites** | | | | |
| Males | | 67 | 66 | 25 |
| Brazil (BSA) | 60 | - | **0.009** | **0.010** |
| Colombia (BSG) | 24 | **0.010** | - | 0.0002 |
| Antarctic Peninsula | 40 | **0.006** | 0.006 | - |

Differentiation indices for males and females are below and above the diagonal, respectively. Significant values are highlighted in bold ($P < 0.001$).

Table 7. Results of potential migrants and individuals with admixed ancestry identified by STRUCTURE and GENECLASS.

| | | mtDNA haplotype | Region | | STRUCTURE | GENECLASS | | |
|---|---|---|---|---|---|---|---|---|
| Individuals | Sex | original/matching | Sampled | Assigned | q-values (BSA/BSG) | $L = L_{home}/L_{max}$ | P value | [log(L)] (BSA/BSG) |
| BR205_02 | Male | ? | BSA | BSG | 0.539/0.461 | 1.374 | 0.0065* | 20.443/19.069 |
| **BR386_06** | Male | ? | BSA | BSG | 0.075/**0.925***** | 3.163 | 0.0002* | 22.342/**19.179** |
| **BR670_10** | Male | ? | BSA | BSG | 0.282/**0.718**** | 2.222 | 0.0020* | 24.071/**21.849** |
| **Mno92Co020** | ? | SP62/HBA104 | BSG | BSA | **0.793**/0.207** | 4.003 | 0.0002* | **21.652**/25.685 |
| Mno93Co009 | Male | SP25/HBA022 | BSG | ? | 0.453/0.547 | 1.163 | 0.0444 | 22.378/23.541 |
| Mno96Co032 | Male | SP14 | BSG | BSA | 0.684/0.316* | 2.005 | 0.0106 | 21.041/23.095 |
| Mno99Co004 | ? | SP62/HBA104 | BSG | BSA | 0.919/0.081*** | 1.749 | 0.0196 | 23.196/24.946 |
| Mno96AP007 | Male | ? | AP | ? | 0.448/0.552 | 0.364 | 0.1163 | 24.323/24.687 |
| Mno96AP019 | Male | SP52 | AP | BSA | 0.585/0.415* | 1.577 | 0.0245 | 19.882/21.399 |

Six individuals identified as potential migrants (***) or admixed (* or **) in STRUCTURE. Individuals identified as F1 (*) migrants in GENECLASS. Individuals identified as migrants with both methods and the most likely source population for each individual is shown in bold. Significant P value ($P < 0.01$) to detect F1 migrants. See Fig. 2 for membership coefficients generated by STRUCTURE.

# Capítulo 4 - The demographic history of the Southwestern Atlantic humpback whales (*Megaptera novaeangliae*) inferred from multiple nuclear loci suggests that the population was declining before whaling

Ana Lúcia Cypriano-Souza, Beth Slikas, Márcia H. Engel, C. Scott Baker, Sandro L. Bonatto

**(Artigo a ser submetido à revista científica *PNAS*)**


"We saw also abundance of large whales, there being more in those southern seas, as I may say, by a hundred to one; then we have to the northward of us."

— William Ambrosia Cowley, *Cowley's Voyage Round the Globe*

# The demographic history of the Southwestern Atlantic humpback whales (*Megaptera novaeangliae*) inferred from multiple nuclear loci suggests that the population was declining before whaling

Ana Lúcia Cypriano-Souza[1,2]; Beth Slikas[3]; Márcia H. Engel[2]; C. Scott Baker[3,4] and Sandro L. Bonatto[1]

[1] *Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga 6681, 90619-900 Porto Alegre, RS, Brazil*

[2] *Projeto Baleia Jubarte/Instituto Baleia Jubarte, Rua Barão do Rio Branco 26, 45900-000 Caravelas, BA, Brazil*

[3] *Marine Mammal Institute, Department of Fisheries and Wildlife, Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365, USA*

[4] *School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1010, New Zealand*

Address for correspondence: Ana Lúcia Cypriano-Souza, Faculdade de Biociências, PUCRS, Av. Ipiranga 6681, 90619-900 Porto Alegre, RS, Brazil, Phone: +55 (51) 3353.4727, Fax: +55 (51) 3320.3612, Email: anacypriano@yahoo.com.br

Running title: demographic history of the humpback whales from breeding stock A based on multiple nuclear loci

**Abstract**

Commercial whaling mainly during the 20th century reduced most populations of the humpback whale. However, the pre-whaling abundances for most of these populations are unknown and there is no knowledge about the historical population dynamics and ecology before the onset of anthropogenic exploitation. Southwestern Atlantic humpback whale population (Breeding Stock A) was estimated by catch records to having reduced to nearly 2% of its historical size by the commercial whaling. Here we inferred the demographic history of this humpback whale population using the 454 sequencing technology to sequence multiple nuclear loci in 24 individuals sampled off the Brazilian coast as well as eight samples from Alaska. Our genetic-based estimate of long-term population size of 147,189 individuals (95% CI: 88,886 - 256,124) was approximately 6 fold higher than the BSA pre-whaling abundance estimate based on catch records of 24,700, which is broadly compatible with the large historical abundance estimate for the humpback whales from North Atlantic Ocean. However, this estimate may probably include the whole South Atlantic metapopulation. Importantly, an extended Bayesian skyline plot (EBSP) analysis indicated that the population was declining ever since a size peak around 30 kya. However, this analysis also detected a decline coincident with the anthropogenic population depletion over past 200 years. Overall, our results found evidence, for the first time, that Southwestern Atlantic humpback whale population was much higher before the onset of the whaling period, which may explain the discrepancy found between previous genetic and catch record population size estimates at this period. This study highlights the importance for taking into account to the historical abundance and environmental impacts for the conservation management plans of a threatened species, and the need for future investigations of the effects of humpback whale population structure in the South Atlantic.

*Key words*: humpback whale, nuclear loci, long-term population size, bottleneck, population dynamics

**Introduction**

Commercial whaling during the 19th and 20th centuries resulted in significant population declines of many baleen whale species, driving some populations to endangerment due to anthropogenic over-exploitation. In the Southern Hemisphere more than 2,000,000 large whales were hunted, including about 200,000 humpback whales (*Megaptera novaeangliae* Borowski, 1781) that were taken after 1900 mainly by whaling operations around Antarctica feeding areas, reducing some populations to small fractions of their pre-exploitation levels (Allison 2010, Clapham and Baker 2002, Findlay 2001). However, the pre-whaling abundances for most of these populations are unknown hindering the exact magnitude of the degree of reduction caused by exploitation. Besides there is no knowledge of the impact of whaling on the features of population ecology, including levels of fidelity to breeding and feeding grounds, migration patterns, and consequently population structure.

The Southwestern Atlantic humpback whale population, recognized as the Breeding Stock A (BSA) by International Whaling Commission (IWC), winters along the eastern and northeastern coast of Brazil (~ 5° to 23° S) (Andriolo et al. 2006, 2010, Martins et al. 2001, Zerbini et al. 2004), and migrates to summer feeding areas around South Georgia and South Sandwich islands in the Southern Ocean (Engel et al. 2008, Engel and Martin 2009, Stevick et al. 2006, Zerbini et al. 2006a, 2011). Since the 17th century, Brazilian humpback whales began to be hunted by whaling operations, which were coastal and of small scale (~ 50 whales per year in Caravelas, Bahia) until the 19th century (Ellis 1969, Lodi 1992). However, modern commercial whaling began in 1904 and expanded the activities to high-density areas in the Antarctic and Sub-Antarctic feeding grounds, which increased the annual catch to several thousand whales and drove the population to the collapse (Findlay 200,1Tønnessen and Johnsen 1982). In the South Georgia island was established the first commercial whaling station in the Southern Ocean, and between 1904 and 1913 about 19,000 humpback whales were killed in the surroundings of the island (Headland 1984).

Although the international protection of the humpback whales from whaling have been established in 1966 (Rice 1978), the former Soviet Union fleet continued taking whales illegally off the central coast of Brazil until 1973 (Yablokov et al. 1998). The BSA pre-exploitation abundance was estimated by historic catch records to nearly 24,700 individuals and reached its lowest numbers in the late 1950s, when there were less than 500 individuals, nearly 2% of its historical size (Zerbini et al. 2006b). Despite this significant population decline, the population has shown signs of recovery with the current rate of increase estimated at 7.4% per annum (Ward et al. 2011) and the most recent abundance in 2008 estimated at

about 9,000 individuals (Wedekin et al. 2010). Recently, the species was reclassified from "Vulnerable" to "Least Concern" by the International Union for the Conservation of Nature and Natural Resources (IUCN 2008). Notwithstanding that these census size changes estimated by whaling catch records and the apparent recovery in the BSA, no study has been done to estimate pre-whaling abundance estimate using coalescent methods to evaluating the genetic signal for the anthropogenic population depletion.

Importantly, evaluating the recovery of whale populations that suffered over-exploitation requires knowledge of historic population sizes and reconstruction of the historical trajectory of decline and recovery (if occurred) (Baker and Clapham 2004, Jackson et al. 2008). The dimension of population recovery has important implications for the future management plans and for understanding the ecological role of whales. Nevertheless, robust estimates of historical abundance are challenging to be obtained. The standard approach to estimating pre-whaling abundance uses a combination of catch records from whaling logbooks and current abundance estimates, which are corrected downward accounting for the rates of increase, and population structure, applying a population dynamic model to estimate changes in population size through time (Baker and Clapham 2004, Jackson et al. 2008). However, these abundance estimates of historical data may be biased downward due to whaling records lost, intentional under-reporting, or inaccurate records (such as struck-but-lost rate) (Baker and Clapham 2004, Clapham et al. 2005).

Particularly, historical population size estimation can be obtained from the level of genetic diversity in contemporary populations (Philips et al. 2012, Palsbøll et al. 2013). The relationship between genetic diversity ($\theta$) and effective population size ($N_e$) is given by the formula $\theta = 4 N_e \mu$, where $\mu$ is the average mutation rate per generation. Genetic diversity is reduced when a population declines, but the reduction in $\theta$ is not as fast as the reduction in census population size ($N_c$) (except in a severe bottleneck) because the processes that alter the genetic diversity, like random genetic drift, mutation and selection, change more slowly than demography. Indeed, the genetic diversity can be relatively unaffected by moderate short-term changes in census population size. Although a drastic population size reduction or bottleneck can leave signatures in contemporary genetic diversity which reflect the demographic history of the population, demographic studies are challenging since the history of each population includes multiple events temporally stratified and of diverse magnitudes (Palsbøll et al. 2013). Accurate long-term population sizes may be difficult to estimate with a single locus (due to the stochasticity of the genetic drift process), as well as due to uncertainty about the mutation rate and generation time used (Clapham et al. 2005, Palsbøll et al. 2013). In addition,

unaccounted population subdivision and migration with other subpopulations also affect the historical size estimates. For example, increasing migration among subpopulations may leave both the sub-population and meta-population $N_e$ estimates equivalent (Waples 2010). Hence, it is important to include multiple unlinked loci, and estimating the locus-specific mutation rates, which increases the accuracy of estimates of $\theta$. Also, it is necessary to determine the levels of population structure (Clapham et al. 2005, Palsbøll et al. 2013).

Recent historical abundance estimates based on genetic diversity have been reported for some baleen whale species that were target of the commercial whaling (Alter et al. 2007, 2012, Roman and Palumbi 2003, Rooney et al. 1999, 2001, Ruegg et al. 2010, 2013, Waldick et al. 2002). Many of these genetic estimates of whale historical population sizes have been noteworthy higher than abundance estimates from catch records (Punt et al. 2006, Wade and Perryman 2002). Many factors may explain this difference between the estimates. First, the abundance estimates based on whaling records may be too low due to incomplete data. In contrast, the abundances estimated from genetic diversity could be too high if very low mutation rates or few genetic markers are used, or if generation time is underestimated, and/or population structure is not accounted for (Baker and Clapham 2004, Clapham et al. 2005). On the other hand, recent long-term abundance estimates in gray whales, *Eschrichtius robustus* (Alter et al. 2007), Antarctic minke whales, *Balaenoptera bonaerensis* (Ruegg et al. 2010), and humpback whales (Ruegg et al. 2013) have included some of the methodological improvements suggested by several authors (Alter and Palumbi 2009, Clapham et al. 2005, Jackson et al. 2009, Palsbøll et al. 2013), reducing some of these uncertainties in estimates. On the other hand, as genetic data provide long-term mean estimate rather than the abundance estimate for the population at a specific time point, it is possible that the population sizes of whales before whaling have been lower than their long-term sizes (Alter et al. 2007, 2012, Ruegg et al. 2013, Palsbøll et al. 2013). In this way, the genetic and catch record inferences may be both accurate. However, this hypothesis has not been tested so far since no study have estimated the size of whale populations before the onset of whaling. Thus, an integration of genetic and demographic approaches, and improvements in estimates are required to a better understanding of the dynamic of whale populations (Baker and Clapham 2004).

As a way of improvement in long-term population effective size estimates and comparing with mtDNA and microsatellite data, an increasing number of studies have employed information of multiple sequence loci (Alter et al. 2007, Ruegg et al. 2010, 2013). Recent studies have used a target gene method to detect single nucleotide polymorphisms (SNPs) and nuclear intron sequences (Lyons et al. 1997, Morin et al. 2007, 2010, Palumbi and

Baker 1994). This approach is based on design of PCR primers, known as exon-primed intron crossing (EPIC) or comparative anchor-tagged sequence (CATS), in ortholog nuclear genes from multiple species to amplify a less conserved region (e.g. microsatellite or intron). In addition, random locus approach, which is based on design of primers in unknown DNA sequences from a previously generated DNA library, is also used to discover SNPs (Morin et al. 2004, 2010). Non-coding introns have many advantages compared to other markers because they are considered neutral, accumulate informative mutations across sites more uniformly, with less homoplasy, and lower transition/transversion rates than coding loci (Palumbi and Baker 1994, Prychitko et al. 2003). Single nucleotide polymorphisms (SNPs) are the most common genetic variation within genomes (coding and non-coding regions), and are characterized by a lower mutation rate and simple mutation model (Aitken et al. 2004, Morin et al. 2004, 2007, Morin and McCarthy 2007). Although SNPs are typically bi-allelic and provide lower genetic power for population analysis than that of microsatellite loci, they can be genotyped using a large variety of technologies which can increase the number of SNPs at several times, increasing their statistical power (Mesnick et al. 2011, Morin et al. 2010). In addition, SNPs can be genotyped in degraded and poor-quality samples, allowing a expansion of the use of historical samples (Morin and McCarthy 2007). Therefore, SNPs represent a more stable nuclear marker becoming an ideal marker for genetic diversity studies. As result, both markers (introns and SNPs) have been previously used as a source of variation for many cetacean studies: SNPs discovery in sperm (Morin et al. 2007), bowhead (Morin et al. 2010), and humpback whales (Polanowski et al.); population structure of humpback (Palumbi and Baker 1994), sperm (Mesnick et al. 2011), and bowhead whales (Morin et al. 2012); and in demographic analysis of gray (Alter et al. 2007), Antarctic minke (Ruegg et al. 2010), and humpback whales (Ruegg et al. 2013).

As increasing the amount of genetic data increases the precision and reliance of the historical abundance estimates (Carling and Brumfield 2007, Takahata 1986, 1995, Yang 1997), the promising technology of next-generation sequencing allows the generation of massive amounts of sequence data by a fast and cost effective way (Descamps and Campbell 2010, Glenn 2011, Harismendy et al. 2009, Liu et al. 2012, Mardis 2008, Morozova and Marra 2008). The 454 pyrosequencing method (Margulies et al. 2005) has been used to generate thousands of target loci reads of around 400 bp per read and to sequence simultaneously pooled PCR products of multiple loci and individuals, which after are analyzed separately (Binladen et al. 2007). Several bioinformatics pipelines have been developed to allow the analysis of hundreds of thousands of sequences, varying from raw data

processing to obtaining genotypes from data (Cole et al. 2009, Goecks et al. 2010, Hird et al. 2011, 2012, Kumar et al. 2011, Stuglick et al. 2011).

Despite three previous studies have estimated the historical abundance for the humpback whales in the North Atlantic from contemporary levels of genetic diversity (Roman and Palumbi 2003, Alter and Palumbi 2009, Ruegg et al. 2013), no study has been made to estimate the genetic-based historical abundance for the other humpback whale populations (or stocks). In this study, we use for the first time the 454 sequencing technology to sequence simultaneously multiple independent nuclear loci in 24 humpback whale individuals, sampled off the coast of Brazil, in combination with coalescent-based methods to infer the demographic history of the Brazilian humpback whale population. In addition, eight individuals from Glacier Bay, in the southeastern Alaska, were included as an outside population to comparison. Initially, we designed 24 primer pairs based on anonymous loci available from RAD tagging of a fin whale, which were added to the nuclear loci set increasing the amount of the final data analyzed. Then we examined all nuclear loci for polymorphic sites (or potential SNPs) and we also compared the results with traditional Sanger sequencing method. Finally, to investigate the demographic history of the population we estimated long-term effective and census population size. Additionally, we use coalescent-based methods to estimate divergence time between the two populations, and also to evaluate past population dynamics. This study is meant to serve both as a test case for the 454 sequencing approach for the species and to increase our understanding of the evolutionary history of the humpback whales from Southwestern Atlantic Ocean.

## Materials and methods

*Sample collection and DNA extraction*

Skin samples from 24 individual humpback whales were obtained by the biopsy dart procedure (Lambertsen 1987) during the winter breeding seasons (July - November) in 2006 - 2011 from breeding ground off Brazil. Samples were preserved in 70% ethanol and were stored at -20°C prior to DNA extraction. Total genomic DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) according to manufacturer's protocol. In addition, DNA samples from eight individuals from Glacier Bay (GB), Icy Strait subregion, in the southeastern Alaska, were obtained through the project Structure of Populations, Levels of Abundance, and Status of Humpbacks (SPLASH). Based on previous microsatellite analysis (Cypriano-Souza et al. in prep) or photo-identification data (CS Baker, personal communication) no duplicate sample was included in this study.

*Primer design*

Twenty three nuclear loci (Table S1) that were previously used for right whales, *Eubalaena sp.*, (Slikas et al. in prep) were selected. From these 23, 18 loci were previously published conserved primer pairs for mammalian and cetacean introns (Lyons et al. 1997, Morin et al. 2010, Palumbi and Baker 1994) and five were primer pairs designed from nuclear DNA sequences from a fin whale generated through RAD-Seq approach in the Biota Science. Further, to increase the number of nuclear loci 34 more primer pairs were designed from approximately 7,000 fin whale DNA sequences generated by RAD-Seq. From 34 primer pairs designed, 24 provided clean sequences with an average length of about 450 bp and were added to the nuclear loci set, totaling 47 loci (Table S1).

A 454 GS Junior sequencer (Roche), available at the Hatfield Marine Science Center (HMSC) of Oregon State University (OSU) in Newport, OR, was used to sequence the 47 PCR amplified nuclear loci of 32 humpback whale individuals (24 Brazilian and 8 Southeastern Alaska samples). As this study includes several samples for multiple amplicons, a universal tailed amplicon sequencing design was used for the generation of the 454 amplicon libraries. For the library preparation two sets of fusion primer pairs were required for a two-step PCR protocol. First, the universal tail M13 and T7 sequences (Univ-A and Univ-B) were added to the 5' end of all 47 forward and reverse template specific primers, respectively. At last, the fusion primer pairs that were used for the second round of PCR were composed of three parts fused together in the 5' to 3' direction: the Primer A or B (forward and reverse, respectively), ending with the sequencing key (TCAG), for binding to the DNA Capture Beads (Lib-A), followed for the multiplex identifiers (MIDs), and the complement primer of the universal tails (Univ-A or Univ-B). A total of 79 primer pairs were synthesized by the Integrated DNA Technologies, Inc. (IDT), of which 47 were target-specific primers for the first round PCR (Table S1) and 32 were barcoding primers for the second round PCR (Table S2).

*454 library preparation, emPCR and sequencing*

First round PCRs were performed in 20 µl final volume with the following reagent concentrations: 0.5 mM of $MgCl_2$, 0.4 mM of dNTPs (Promega, USA), 0.5 µM of each primer (IDT), 1 mg/ml of BSA (Bovine Serum Albumin), 2% of DMSO (NEB, USA), 1 X Phusion HF buffer (NEB, USA), 0.3 units of Phusion High-Fidelity DNA Polymerase (NEB, USA), 9.45 µl of $ddH_2O$, and 1 µl of template DNA (~ 20 ng). Thermocycle profile consisted of a

touch-down with an initial denaturation at 98°C for 30 s followed by 3 cycles of denaturation at 98°C for 8 s, annealing at 60°C for 30 s, and extension at 72°C for 1 min 15 s; followed by 3 cycles of denaturation at 98°C for 8 s, annealing at 58°C for 30 s, and extension at 72°C for 1 min 15 s; followed by 30 cycles of denaturation at 98°C for 8 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min 15 s, and then by a final extension at 72°C for 15 min. All PCR products were checked by UV imaging of agarose gels for successful amplification. Reactions were purified with 1.8x AMPure XP beads (Agencourt) using the manufacturer's protocol.

For the second round PCRs, 1 µl of the AMPure-cleaned 1$^{st}$ round PCR product was amplified in a 50 µl final volume with the following concentrations: 0.5 mM of $MgCl_2$, 0.2 mM of dNTPs (Promega, USA), 0.5 µM of each primer (IDT), 2% of DMSO (NEB, USA), 1 X Phusion HF buffer (NEB, USA), 0.4 units of Phusion High-Fidelity DNA Polymerase (NEB, USA), and 31.3 µl of $ddH_2O$. Thermal cycling conditions were an initial denaturation at 98°C for 30 s, followed by 12 cycles of denaturation at 98°C for 8 s, annealing at 60°C for 30 s, extension at 72°C for 45 s, and a final extension at 72°C for 15 min. Two steps of pooling and purification using 0.6x AMPure XP (Agencourt) were performed for each locus. First, four individual 2$^{nd}$ round PCR products were pooled for each locus and purified, followed by the pooling of all samples for each locus and purification. PCR products were then quantified using a Quant-iT$^{TM}$ PicoGreen dsDNA (Invitrogen) and pooled to obtain 2 libraries (LB1 = 23 loci, and LB2 = 24 loci). Each library was purified twice using the AMPure XP beads with a 0.7 bead to DNA ratio. Quality control PCR (QC PCR) was performed for each library to check for the presence of short fragments. The libraries were quantified using the Quant-iT$^{TM}$ PicoGreen dsDNA (Invitrogen) and diluted according to the manual's recommended concentration.

Emulsion PCR (emPCR) was performed as described in the GS Junior Titanium emPCR Amplification Method Manual - Lib-A (March 2012 version - Roche) using 0.5 molecules per bead for each library and the GS Junior Titanium emPCR Kit (Lib-A). After emPCR, the beads with the amplified DNA library were recovered and washed, followed by enrichment of the DNA-carrying beads and annealing of the sequencing primers. Finally, enriched beads were counted with a GS Junior Bead Counter v2 to evaluate the amount of beads recommended (500,000 beads) to the sequencing. Each enriched library was sequenced in a 454 GS Junior sequencer (Roche), according to the Sequencing Method Manual using the GS Junior Titanium Sequencing Kit and the Pico-Titer plate Kit (Roche). Sequencing was started with the "Full processing for Amplicons" which is appropriate for Amplicon libraries.

*454 data processing and variable sites detection*

The raw data (.fna and .qual files) generated by the 454 sequencer was first processed in the Ribosomal Database Project (RDP) Web site (Cole et al. 2009) which has a Pyrosequencing Pipeline that separated the reads by barcodes (MIDs) and performed the quality control (QC) of the reads, filtering the high quality (Q > 20 and fragment length > 150 bp) and eliminating the low quality reads. The second step was the assembly of high-quality reads to a provisional reference genome (PRG) using the open source pipeline PRGmatic (Hird et al. 2011), which constructs its own reference genome from the loci targeted using dependent programs ($CAP_3$, BWA, SAMTOOLS and VARSCAN). In brief, PRGmatic used as input the reads sorted by MID tag (.fasta and .qual files) generated by RDP, and clustered the reads at high similarity (99%) identifying putative loci within each individual using $CAP_3$ (Huang and Madan 1999). A consensus sequence for each putative locus was generated by high coverage (default = 5x) alleles within each individual. These putative alleles were clustered at lower per cent identity (90%) across all individuals into provisional loci, which were concatenated to form the PRG. Then all original reads were aligned to the PRG using BWA (Li and Durbin 2009) and SAMTOOLS (Li et al. 2009) to output a summary of each individual's reads in pileup format. Individuals at each locus with a minimum depth of coverage (default = 6x) were retained. In the final step, PRGmatic used VARSCAN (Koboldt et al. 2009) and custom Perl scripts to call two alleles per individual using the minimum coverage cut-off value of reads to call a variable site or SNP at a particular base position (default = 3x) and the minimum incidence of a SNP variant required to call a heterozygote individual (default = 20%).

To further improve the detection of all loci and potential alleles a manual review was done as following. All contigs (putative loci) with all reads trimmed generated by BWA (SAM file) for each individual were imported into Geneious 5.6.6 (Biomatters, Auckland, NZ) to call genotypes. Considering the known probability of NGS errors generated during PCR and pyrosequencing, it is crucial to detect and discard the reads with sequencing errors or that correspond to other loci than the targeted locus. To accomplish this step, the resulting contigs were aligned to the PRG, built based on known nuclear loci used to design the primers, using Muscle (Edgar 2004) in Geneious to identify the targeted loci. In this way, the contigs that did not map to the PRG were reviewed and blasted against sequences available in GenBank. The sequences not identified were considered potential recombinant chimeric sequences or pseudogenes generated by non-specific amplification, and were discarded. For

the remaining sequences identified as targeted loci a validation process was performed. The first step was removal of reads with incomplete specific primers or MID tags. The next step was to discard the samples with a low number of reads which may induce an incomplete genotyping. The last step was to remove the SNPs represented by a low number of reads for a given sample since these variants represented only rarely within samples probably resulted from sequencing errors.

Calling genotype for each individual was performed for the contigs (loci) that map to the PRG and passed through the validation process. First, all reads of all individuals were aligned for each locus to identification of the possible polymorphic sites. Then the reads of each individual was reviewed for each locus, visualizing the variations and errors, and defining the alleles. To confirm the reliability of these alleles three main filtering criteria were required for each polymorphic site: 1) coverage by both forward and reverse reads (attending to the problem of strand bias), or 2) at least five reads (5x coverage) with quality score over 20, or 3) if heterozygous, the minor allele frequency was > 0.15 and at least one read from each allele had the quality score over 20. As the error rate for homopolymers (extensive sequence of the same base pair) in 454 is non-negligible, if the variation was an indel in this region, the criteria used was the majority consensus of all reads. Also, in order to further validate the 454 sequencing, mainly the homopolymeric regions, two individuals (one from Brazil and other from Southwestern Alaska) were sequenced for each locus on a Sanger sequencer (ABI 3730XL, Applied Biosystems) using the same PCR used for the 454 sequencing.

Once all individual allelic states were detected for each locus, they were aligned again for a last checking. Most singleton alleles (alleles found in only one individual) had high coverage (with many reads), but three singletons (in the BH42b, BH395 and BH404 loci) occurred only in few reads (2 to 4) and thus were not included in the analyses, except one (in the BH404 locus) that was later confirmed by Sanger sequencing. Heterozygous base pairs in the sequence sorted two different haplotypes which were considered as the two allelic phase for the locus, attending to the problem of phase nuclear data. Some reads were removed of the heterozygous locus, assuming to be artifacts of PCR, emPCR or sequencing since they occurred in small number and appeared to be recombinants between the two allelic states. All variants within primer regions were masked. Finally, the variable sites validated for each locus were placed into an Excel spreadsheet taking into account the number of reads for each allele and individual.

*Data analyses*

*Genetic diversity and testing for population structure*

Simple statistics for the whole dataset were calculated for the total number of reads, the total number of samples, the total number of alleles, and the total number of variable sites for each locus. Genetic diversity was estimated as haplotype/allele richness (A) and nucleotide diversity ($\pi$) for each locus in each population using DnaSP v.5.0 (Librado and Rosas 2009). Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997) neutrality tests were also calculated for each locus in each population using DnaSP to evaluate signal of demographic expansion or contraction (significantly negative or positive values, respectively). The minimum number of recombination events in the sample was also calculated using the same program.

Although the two humpback whale populations (BSA and GB) of this study are located in two different Oceans (Southern Atlantic and Northern Pacific) and Hemispheres (Southern and Northern Hemispheres), population structure was tested to assess the potential of detection of population differentiation within a multi-locus framework. A Bayesian clustering approach for inferring population structure was performed in STRUCTURE v. 2.3.3 (Pritchard et al. 2000, 2007). We used the approach in which the haplotypes at each locus were recoded as alleles, as suggested previously for a similar data set (Ruegg et al. 2013). Ten independent runs were performed for each $K$ between 1 and 5 with no prior information on sampling location using the admixture and correlated allele frequencies model. Burn-in and length of simulation were set at 500,000 and 1,000,000 iterations, respectively. In the next analysis the sampling location prior was also fitted to assist the identification of clustering for a weak population structure (Hubisz et al. 2009). The results generated were processed in STRUCTURE HARVESTER v. 0.6.93 (Earl and vonHoldt 2012), a web-based program which plots the mean of the likelihood values per $K$ and the ad hoc $\Delta K$ (Evano et al. 2005) to estimate the most likely value of $K$.

*Demographic history*

First, we tried to estimate the long-term effective population size ($N_e$) of these two samples taking into account a possible gene flow between the populations. Based on this, two coalescent-based methods implemented both in *lamarc* version 2.1.8 (Kuhner 2006) and *migrate-n* version 3.4.2 (Beerli 2006, 2008) were employed to estimate $\Theta$ ($\Theta = 4N_e\mu$, where $\mu$ is the mutation rate per generation) and migration rate ($M = m/\mu$, where $m$ is the probability of immigration per generation per gene copy) for each population. However, it was not possible

to stabilize the estimates for migration rates after long runs with various parameters options (data not shown). Therefore, the following analyses have focused on estimating just the effective population sizes for each population. *Lamarc* analysis used a Bayesian approach for three independent Markov Chain Monte Carlo (MCMC) runs of one long chain performed with 5,000,000 generations, sampling every 500$^{th}$ genealogy estimate after discarding the first 1,000 generations. For the analysis in *migrate-n*, a maximum likelihood (ML) approach was used and three replicates searches of 20 short chains with a total of 100,000 genealogies and two long chains with 1,000,000 genealogies, with a sampling increment of 100 steps and a burn-in of 10,000 steps. *Migrate* was run three times to ensure consistency between runs. Tracer 1.5 (Rambaut and Drummond 2007) was used to assess the convergence of the coalescent analyses by inspecting the effective sample sizes (ESS) and the stationarity of posterior distributions.

For the conversion of $\theta$ into long-term effective population size it was used the previously published average mutation rate of 4.4 x $10^{-10}$ per base pair per year for nuclear loci of the humpback whales from Ruegg et al. (2013). This mutation rate was converted in units of mutation per base pair per generation using the generation time of 18 years taking into account the range between 12 and 24 years estimated for the North Atlantic humpback whales (Chittleborough 1965, Roman and Palumbi 2003). The total census population size ($N_c$) was calculated from the effective size ($N_e$) by multiplying by a conservative 2:1 ratio of total mature adults to the effective number of adults ($N_{mature}$: $N_e$) (Alter et al. 2007, Roman and Palumbi 2003, Ruegg et al. 2010, 2013), and by the proportion of juveniles in the population (number of adults + juveniles)/(number of adults), estimated between 1.6 to 2.0 for humpback whales (Chittleborough 1965, Roman and Palumbi 2003). Therefore, the average ratio of census population size to effective population size estimated was 3.6, with a variation from 3.2 to 4.0.

Additionally, we also used the isolation with migration model implemented in the program IMa2 with a MCMC procedure within a Bayesian inference, and that incorporates parameters for time since divergence between the populations, bidirectional gene flow and the ancestral and the two descendent population sizes (Hey and Nielsen 2004, Hey et al. 2004, Hey 2010, Nielsen and Wakely 2001). Most nuclear loci were fit to the infinite sites (IS) model while only four were fit to the HKY model. A uniform prior distribution was used with maximum priors set at $\theta_{max} = 3$ for all populations (ancestral and descendants), $m_{max} = 3.3$ for both gene flow direction, and $\tau_{max} = 1.2$. Three simulations were run for 10,000,000 genealogies with a sampling interval of 100 genealogies, and a burn-in of 500,000 steps.

Metropolis-coupling was used for 40 coupled chains that varied over a range of heating (ha = 0.96, hb = 0.9) values to achieve adequate mixing. Mutation rate estimates per year per locus (not per base pair) for all loci were included to obtain estimates of effective population sizes, migration rates and the time of divergence (demographic parameters). For time to be scaled properly the generation time in years was specified.

Finally, for the Brazilian humpback whales an extended Bayesian skyline plot (EBSP) method (Heled and Drummond 2008) implemented in the program BEAST 1.7.5 (Drummond and Rambaut 2007) was also used to estimate the population dynamics over time. The EBSP is an extension of the BSP incorporating multiple loci for inferring non-parametric population size changes over time based on the coalescence and accounting for uncertainty using a MCMC procedure. Control region mtDNA haplotypes from 158 individuals available from previous studies (Cypriano-Souza et al. in prep, Engel et al. 2008) were added to the 35 nuclear loci. A normal distribution for both (mtDNA and nuclear loci) substitution rate priors with a strict clock model was used in the analysis. For the mtDNA control region sequences we used the mean substitution rate of $3.9 \times 10^{-8}$ (s.d. $1.5 \times 10^{-8}$) with a HKY with 6 gamma categories model. For the nuclear loci we used a mean substitution rate of $4.4 \times 10^{-10}$ (s.d. $1.8 \times 10^{-10}$) with a HKY with no gamma model. The program was run for 500,000,000 generations, with a sampling interval of 50,000 generations, and with the first 10% of generations discarded as burn-in. Results were checked for ESS of more than 200 and convergence of posterior values using Tracer.

**Results**

*454 quality control, coverage, and validation of variable sites*

A total of 130,712 reads were generated for two libraries (LB1 = 50,480 reads and LB2 = 80,232 reads) through three GS Junior sequencing runs (one run for the LB1 and two runs for the LB2). After quality filtering and removal of reads lower to 150 bp, 96,678 (73.9%) high-quality reads (LB1 = 40,171 reads, and LB2 = 56,507 reads, respectively 79.6% and 70.4% of the total generated for each library) were used for the loci identification analysis, and for the variable site validation process (correlated to the coverage of the locus and of the variation). Of these, 87,653 (90,6% of the high-quality reads) reads with an average read length of 440 bp were identified as the targeted loci and subsequently used for the next analyses. Importantly, the BLASTn search employed against all sequences in GenBank for each putative locus found by PRGmatic have confirmed all known loci (18 previously published nuclear introns (Lyons et al. 1997, Morin et al. 2010, Palumbi and Baker 1994)).

However, for the CYP1A1 locus that was split in two contigs, the BLASTn search found two different genes, the targeted amplicon CYP1A1, and the paralog gene CYP1A2 (Niimi et al. 2005). Total reads analyzed for the libraries LB1 and LB2 were 36,791 (91.5%) and 50,862 (90%), respectively, representing 22 different loci for each library, and totaling 44 loci sequenced (Table 1). This difference in reads number between the libraries was probably attributed to the difference in DNA concentration of each library purified, which was 53.40 ng/µl and 159.51 ng/µl for LB1 and LB2, respectively. Although the LB2 provided more reads, four loci were under-represented (relative coverage < 0.10) while only one locus in the LB1 obtained low reads coverage (relative coverage < 0.09). This may be due probably to the pooling of the PCR products (amplicons) with similar concentrations for the first library. These loci were discarded of the analyses (Table S3).

Finally, 86,390 reads (LB1 = 36,374 reads, and LB2 = 50,016 reads) were successfully validated for 39 nuclear loci (LB1 = 21 loci, and LB2 = 18 loci) in 30 individuals of humpback whale, generating over 17 kbp (LB1 = 9,087 bp, and LB2 = 8,055 bp) of DNA sequence per individual (Table 1). The number of reads obtained for each individual depended on the sample considered, with some samples being under-represented or even unrepresented, as the two individuals (BR603-09 and BR718-11) that had no reads in both libraries, and another individual (BR769-11) that also had no reads in the LB1. This may be explained by the poor DNA quality or low amplicon concentrations for these samples before pooling. Overall, a total average of 74 reads per locus per individual (ranging from 1 to 343) was obtained, with an average of 60 and 93 for LB1 and LB2, respectively. Despite the large number of discarded reads (33%) during the loci identification analysis and validation process, the average coverage was about 2,215 reads (LB1 = 1732x, and LB2 = 2779x) per locus, ranging from 241 to 6,402 reads per locus (Table S3). Also, over 66% (LB1 = 51%, and LB2 = 82%) of the target fragments (locus per individual) were covered by more than 40 reads, the coverage sufficient to detect heterozygotes with high confidence, with over 95% of them with at least 1x coverage and over 91% with at least 5x coverage.

After alignment of the 39 loci and completion of the filtering criteria for the reliable polymorphic sites, a total of 99 variable sites (or potential SNPs) and 8 indels (LB1 = 56 variations and 2 indels, and LB2 = 43 variations and 6 indels) were identified among 35 loci for a total of 17,142 bp of DNA sequence per individual. This have resulted in an average SNP frequency of one SNP every 173 bp. However, polymorphism distribution was not uniform across all loci, with five loci containing multiple variable sites (between 5 and 8). These variable sites have characterized 133 different alleles (LB1 = 66 alleles, and LB2 = 67

alleles) for 39 loci and 30 individuals, ranging from 1 to 8 alleles per locus (Table 2). Four loci with high coverage (average coverage of ~ 2760x per locus) were monomorphics. Importantly, many of the polymorphic sites were confirmed with the sequences from two individuals sequenced by Sanger sequencing. These two individuals, sequenced initially by 454 sequencing, showed a total of 67 alleles (50%) of the 133 identified for all loci (generated by 454), which included the single alleles of four monomorphic loci and 13 homozygous alleles. The remaining 50 alleles (heterozygous) were defined by 43 variable sites (62%) of the 69 identified for the other 22 loci for all individuals. Sanger sequencing confirmed 29 (67.5%) of these 43 sites, and thus validating 58 (86%) of 67 alleles, including two quite different alleles with 6 variable sites for two different loci (2 alleles of the ACTIN locus and 2 alleles of the BH395 locus). Also, one singleton allele (found in the BH404 locus) differentiated by one polymorphic site and represented by only four reads was confirmed by Sanger sequencing. All homopolymers with length varying between 3 and 7 base pairs were consistent between the two sequencing methods. Only one homopolymer (the longest) in the BH96 locus showed 11 base pairs ($T_{11}$) in the Sanger sequences and 10 ($T_{11}$) in the 454 alleles. However, the BH96 sequences generated by Sanger sequencing had poor quality, and thus the length of 10 base pairs for this homopolymer was retained. Also, the under-representation of the polymorphic sites in the Sanger sequences is due mainly to the poor quality of some sequences. In conclusion, 454 sequencing was more sensitive to detecting alleles and heterozygotes than Sanger sequencing method, which failed to detect 14 variable sites and 9 alleles, consequently under-representing heterozygous.

*Genetic diversity and population structure*

Genetic diversity at all nuclear loci was low to moderated (Table 3), with a mean allelic richness of about 4 alleles per locus, ranging from 2 to 8. Nucleotide diversity ($\pi$) among 34 polymorphic loci averaged 0.00132, ranging from 0.00010 (CYP1A1) to 0.00728 (BH395). As may be expected given the low substitution rate, neutrality tests (Tajima's D and Fu's FS) for individual locus showed no significant evidence for non-equilibrium population (Table 3). Although Tajima's D for ACT in GB showed a significantly positive value (2.303, $P < 0.05$), this locus does not remained significant after Bonferroni correction for multiple comparison ($P = 0.05/68$ tests $= 0.0007$). Only 3 of 25 loci with more than 2 polymorphic sites showed some evidence of recombination (Table 3).

Given the high level of allele sharing between the two population, it is not unexpected that the Bayesian clustering analysis implemented in STRUCTURE using the model without

(standard) sampling location prior were not very informative to distinguish the populations as seen in the barplots (Fig. 1a and b), although both the average of the likelihood values per $K$ and $\Delta K$ method of Evano et al. (2005) suggested $K = 2$. The barplots of the analysis with the sampling location prior slightly suggest a distinction between these two populations (Fig. 1c and d), although the statistics were ambiguous. Thus, these results indicate that much more polymorphic loci and/or more samples are needed to obtain reliable results with this type of analysis.

*Demographic history*

Estimates of $\theta$ using the coalescent analyses in *Lamarc* and *Migrate* were slightly different, but the 95% confidence intervals showed overlap between the two estimates. Although we report only the estimates with *Lamarc*, the results from *Migrate* analyses are shown in Table 4. *Lamarc* yielded for all 39 loci (including 4 monomorphic loci) an overall most probable $\theta$ estimate (MPE) of 0.001292 (95% CI = 0.000975 - 0.001620) for the Brazilian (BSA) humpback whales and of 0.000759 (95% CI = 0.000562 - 0.001092) for the Southeastern Alaska (GB) whales (Table 4). For each locus, estimates of $\theta$ varied from 0.00001 to 0.0035 for the BSA and from 0.00001 to 0.0044 for the GB (Table S4) which reflects variation in mutation rate and/or coalescent history among loci. Based upon the mutation rate of $4.40 \times 10^{-10}$ (95% CI: $3.66 \times 10^{-10}$ - $5.29 \times 10^{-10}$) per base pair per year from Ruegg et al. (2013) and a generation time of 18 years (ranging from 12 to 24), the estimates of $\theta$ were converted into a $N_e$ of 40,886 (95% CI: 27,777 - 64,031) for the Brazilian humpback whales and into a $N_e$ of 24,018 (95% CI: 16,011 - 43,162) for the Southeastern Alaska humpbacks (Table 4). For the 95% confidence intervals (CI) of $N_e$ the 95% CI for both $\mu$ and $\theta$, and the range of generation time were incorporated. Long-term effective size was converted into long-term census size ($N_c$) by multiplying by the average ratio of $N_c/N_e$ of 3.6 (ranging from 3.2 to 4), which took into account for the variation in reproductive success (2:1 ratio of total mature adults/breeding adults) and for juvenile abundance (1.6 to 2 ratio of total population size/total adults). Finally, the $N_e$ estimates multiplied by this ratio yielded a long-term $N_c$ of 147,189 (95% CI: 88,886 - 256,124) individuals for the BSA and of 86,464 (95% CI: 51,235 - 172,648) individuals for the GB humpback whales. The confidence intervals incorporated uncertainties in measures of genetic diversity, generation time and juvenile abundance.

In order to investigate the population history of this species, three IMa2 analyses were conducted to estimate the parameters for divergence time, ancestral and descendent

population sizes, and direction of migration between the two populations. All three runs yielded similar results and the Markov chain seems have sampled from a stationary distribution of parameter values since the effective samples sizes were high (most > 10,000) and plots of the sampled parameter values did not show trends. The divergence time ($\tau$) between the two populations (BSA and GB) from a common ancestral population was estimated to have occurred approximately 34,000 years ago. The ancestral $\theta$ was estimated at 0.5025 (95% HPD: 0.3795 - 0.6435) which was converted into a historical $N_e$ of 35,830 (95% HPD: 27,060 - 45,884). The $\theta$ estimated for the two populations resulted in a $N_e$ (58,077) for the BSA larger than either the GB (7,594) or the ancestral (35,830) populations (Table 5), suggesting a historical expansion for the BSA humpback whale population from smaller ancestral population. In contrast, the smaller $N_e$ for GB from ancestral population suggests a historical bottleneck event (Table 5). The high estimates of gene flow from GB to BSA and a too low gene flow in the opposite direction are uninformative since the marginal posterior densities were flat.

For the extended Bayesian skyline plot analysis, the assessment of the effective sample sizes in TRACER have shown values higher than 200 for all parameters, indicating sufficiently deep sampling (Drummond and Rambaut 2007). The results (Fig. 2) suggest that the Brazilian humpback whale population started to expand around 200 kya and peaked about 30 kya, when this population would have reached around 200,000 individuals. This was followed by a continuous population reduction until around 200 years ago, when it reached about one-fourth of the previous maximum size. Very interestingly, this suggests that this population was clearly declining much earlier than the start of the whaling period. Also very importantly, the EBSP indicated a population decline around 200 years ago (~ 11 generations ago) with a stabilization more recently (Fig. 2c). However, it should be noted that the 95% highest posterior density for the most recent times are wide.

## Discussion

*454 sequencing and variable sites validation*

Incorporating information from a large number of non-recombinant loci, containing a range of linked SNPs, is important for accurate estimates of demographic parameters and phylogeographic analyses. This approach has been facilitated recently by the use of NGS technologies that generate a large amount of sequence data at affordable costs for non-model organisms (Gompert et al. 2010, Hohenlohe et al. 2011, McCormarck et al. 2012, Puritz et al. 2012). Despite these technical advances, this is the first study to use tens of multiple unlinked

nuclear loci to investigate the demographic history of the humpback whales, and one of the first to use more than 30 loci in any whale species. We used 454 pyrosequencing and the parallel tagged sequencing (PTS) approach to obtain data for 18 previously published nuclear introns (Lyons et al. 1997, Morin et al. 2010, Palumbi and Baker 1994) as well as for 29 novel anonymous nuclear loci (Slikas et al. in prep, this study) from 30 humpback whale individuals representing two populations, the Southwestern Atlantic Ocean (Breeding stock A) and the Southeastern Alaska (Glacier Bay). This sequencing strategy validated 39 nuclear loci, of which 35 were variable for these populations, generating 99 polymorphic sites (or potential SNPs), and thus being potentially informative for estimating population genetic parameters.

However, our data set showed a difference in coverage across loci and individuals, whereas three loci were not sequenced and five loci were under-represented with low read coverage (relative coverage < 0.10), and three individuals were unrepresented (BR769-11 obtained reads only for the 2$^{nd}$ library). This may be explained by the failure to standardizing the PCR products before the pooling step, with a deficiency in the recovering of haplotypes at particular loci and the poor DNA quality for some samples. One of the problems commonly found at the amplicon sequencing approach is the requirement of standardization of the PCR products to the same concentration to avoid overrepresentation of some loci (Ekblom and Garlindo 2010). Another problem is that some loci could have been affected by short sequences, which were probably the result of incomplete emPCR or pyrosequencing reactions. However, it seems that our amplicons had few short sequences since after quality filtering and removal of short reads (< 150 bp) 73.9% of all reads generated had an average length of about 440 bp, including the five loci under-represented.

Overall, our PTS resulted in little missing data given that individuals showed just 18% of the missing loci. Zellmer et al. (2012) have discussed that although the quality of estimates generated using coalescent models (e.g. BEAST, IMa2) may be associated to the number of alleles, missing data per se should not affect these estimates. Given that the NGS sequencing has several challenges, mainly the identification of paralogous loci and the discrimination of sequencing error from rare variants, it was important to ensure that these errors did not impact our nuclear diversity estimates. Our validation process and comparison with Sanger sequencing have shown that the loci identification and subsequent calling of genotypes were highly accurate. First, we used the PRGmatic pipeline (Hird et al. 2011) that has been used successfully for the processing of NGS data for other non-model organisms (McComarck et al. 2012, Zellmer et al. 2012). On the other hand, our combined parameter thresholds did not allow for calling of all potential alleles with this pipeline. This can be explained by the use of

99% clustering threshold to identify loci, which was too high and consequently separated the more variable alleles into different loci (oversplitting), resulting in few heterozygous individuals. Zellmer et al. (2012) have recommended a clustering level of 95% since this threshold generated the fewest loci with evidence of oversplit or overlumped. For the next steps of validation process and filtering criteria we used Geneious which has also been used efficiently for NGS data analysis in other study (Puritz et al. 2012).

Finally, Sanger sequencing of two individuals have validated many polymorphic sites and homopolymers. In addition, the 7 polymorphic sites identified at 4 alleles of the ACTIN locus generated by 454 sequencing were also confirmed by previously published alleles for this locus in humpback whale samples (Jackson et al. 2009). However, 454 sequencing identified more variable sites and alleles than Sanger sequencing, showing to be more sensitive to detecting genetic variation. Some of the evidences that 454 sequencing is not inflating estimates of genetic variation is the high level of identity (86%) in alleles generated by both sequencing methods, and the four monomorphic loci with high number of reads. In conclusion, the results from this study have demonstrated that 454 parallel tagged sequencing is an accurate method to providing a large number of multiple nuclear loci from multiple individuals in a non-model species. Recently, Polanowski et al. (2011) described 45 TaqMan SNPs markers which will be an alternative genotyping for humpback whales. Our 99 SNPs in conjunction with these 45 SNPs will increase the data set of a marker more suitable for long-term and collaborative studies of a worldwide distributed species.

*Genetic diversity and demographic history*

Overall, levels of genetic variation of the nuclear loci for the Southwestern Atlantic ($\pi$ range: 0.00010 - 0.00728) and the Southeastern Alaska ($\pi$ range: 0.00028 - 0.00602) humpback whales were similar or slightly higher than the nuclear genetic diversity of the Eastern Pacific gray whales ($\pi$ range: 0.00016 - 0.00310; Alter et al. 2007) and of the Northern Atlantic humpback whales ($\pi$ range: 0.00020 - 0.00540; Ruegg et al. 2013). However, our estimates are based on a data set approximately four times larger (35 loci) than those of the previous studies (9 loci). The actin intron, the only nuclear locus common between here and Ruegg et al. (2013), showed higher nucleotide diversity for the Brazilian and Southeastern Alaska humpback whales than Northern Atlantic whales. One possible explanation for the higher nucleotide diversity for this locus in this study may be a higher number of heterozygotes in our data set due to more efficient detection of heterozygotes by 454 sequencing.

Southwestern Atlantic humpback whale population was reduced to nearly 2% of its historical size, to an estimate of 500 individuals in the late 1950s, due to severe exploitation by the commercial whaling (Zerbini et al. 2006b). However, our previous studies have shown no evidence of reduced diversity or a significant genetic bottleneck in this population using mtDNA, microsatellite loci, and standard methods (Cypriano-Souza et al. 2010, Engel et al. 2008). On the other hand, in this study our genetic-based estimate of long-term census population size ($N_c$) of 147,189 individuals (95% CI: 88,886 - 256,124) was much higher (6 times) than the highest BSA pre-whaling abundance based on catch records, estimated at 24,700 individuals (95% CI: 23,016 - 31,795) (Zerbini et al. 2006b). This apparent discrepancy between genetic and catch-based estimates of historical population size may be explained by the different time frames that these estimates represent. While pre-exploitation abundance estimates based on whaling catch data represent population size just prior to the onset of whaling (very recent timescale), the long-term genetic estimates represent the weighted harmonic mean of population sizes over $4N_e$ generations (therefore, thousands of generations), being thus influenced by population demographic dynamics (Charlesworth 2009, Palsbøll et al. 2013). In this way, the genetic and demographic estimates could both be correct if the population sizes of whales before whaling were lower than their average long-term sizes, which seems to be true for most whale population estimates (Alter et al. 2007, 2012, Alter and Palumbi 2009, Ruegg et al. 2013).

Anyway, there are considerable uncertainties for both these estimates that must be taken into account. At first, the reliance of catch records has been questioned due mainly to intentional under-reporting by the former Soviet Union (Yablokov 1994), as well as inaccurate records, such as struck-but-lost rate. All these inaccuracies in the catch records may lead to an underestimated pre-whaling abundance (Baker and Clapham 2004, Clapham et al. 2005, Lubick 2003). In addition, the relationship between current genetic diversity and long-term population size can be complicated due to several factors, such as changes in population size, past hybridization, population structure and departures from random mating. Likewise, there are uncertainties on several parameters surrounding the estimation of long-term population size, including mutation rate, generation time, relation between $N_c$ and $N_e$, and the temporal-scale to which the estimate is applied (Alter et al. 2007, Charlesworth 2009, Clapham et al. 2005, Palsbøll et al. 2013). Consequently, inaccurate long-term population sizes can be estimated if these uncertainties are not taken into account. As several authors have emphasized, evaluating patterns of genetic variation among multiple unlinked loci increases the accuracy of estimates of $\theta$ than using a single locus because each locus

independently assess past population patterns (Carling and Brumfield 2007, Takahata 1986, 1995, Yang 1997).

Therefore, this study have adopted a number of recommendations to a better improvement of long-term population size estimates, including multiple independent nuclear loci, a most reliable nuclear loci mutation rate estimated for the species, taking into account the migration, and the historical time frame of genetic estimates (Alter et al. 2007, Clapham et al. 2005, Palsbøll et al. 2013, Ruegg et al. 2013). Despite the generation time estimates for most whales are based on population mean age, these estimates remain uncertain. Also, generation time may not be stable across evolutionary time scales. In order to take into account for uncertainties of the generation time estimates, we used a wide range (between 12 and 24 years) of estimates for the North Atlantic humpback whales (Chittlebourough 1965, Roman and Palumbi 2003, Taylor et al. 2007). There are other factors that may influence on estimation of $N_e$, such as variance in reproductive success and effects of selection at individual loci that may increase or decrease our long-term population size estimates. However, balancing selection is unlikely since all of the multiple loci showed no departure of neutrality, suggesting that they are evolving according to neutrality and equilibrium.

The point estimates of long-term effective population size based on two coalescent methods were strikingly similar, given they are very complex methods, and their confidence interval overlap widely. The *Lamarc* point $N_e$ estimates were around 41,000 and 24,000 individuals for the Southwestern Atlantic and Southeastern Alaska humpback whales, respectively, while the *Migrate* estimates for Southwestern Atlantic humpbacks was around 41,000 and a little smaller for the Southeastern Alaska whales (~16,000). As our results showed strong convergence among different runs in *Lamarc*, most of our further discussion is based on their results. Importantly, our adjusted effective sizes into long-term census population sizes were similar to previous genetic-based census size estimates for Northern Atlantic humpback whales and for other baleen whale species which have shown much higher genetic-based estimates of historical abundance than those based on catch records (Alter et al. 2007, 2012, Jackson et al. 2008, Roman and Palumbi 2003, Ruegg et al. 2010, 2013). For example, in a first study of gray whales in the eastern North Pacific Ocean, Alter et al. (2007) estimated 96,000 individuals (95 % CI 78,500 - 117,700) based on 9 nuclear introns and cytochrome-b, and after that the authors found a correspondent estimate of 100,670 whales (90% HPD: 59,940 - 111,550) based on mtDNA control region (Alter et al. 2012).

Previous estimates of long-term population size for the humpback whales from North Atlantic Ocean have shown a significant difference, mainly between the original mtDNA-

based estimate of 240,000 individuals (95 % CI: 156,000 - 401,000; Roman and Palumbi 2003), the updated mtDNA-based estimate of 150,000 individuals (95 % CI 45,000 - 180,000; Alter and Palumbi 2009) using a more accurate mutation rate estimate, and the multi-locus estimate of 112,000 individuals (95 % CI 45,000 - 235,000, Ruegg et al. 2013). This difference was due to a lower mutation rate estimate (two times lower) of the mtDNA control region used in the original study. Ruegg et al. (2013) used an average mutation rate across nine nuclear loci according to the phylogenetic analysis of the baleen phylogeny and fossil history (Jackson et al. 2009). Indeed, the mtDNA-based estimate using a recalibrated control region mutation rate is indistinguishable from the multi-locus estimate of historical abundance, highlighting the significance of accurate mutation rates to estimation of long-term population size. Thus, we also used this more reliable nuclear mutation rate for our estimates. Actually, our estimates of long-term population size of ~150,000 individuals (95% CI: ~90,000 - 260,000) and of ~90,000 individuals (95% CI: ~50,000 - 170,000) for the humpback whales from breeding stock A and Glacier Bay (Southeastern Alaska), in the Southwestern Atlantic and North Pacific Oceans, respectively, were more similar to the two more recent estimates for the humpbacks from North Atlantic.

Considering that population structure and migration are parameters that affect the estimated of the genetic diversity within populations, a better estimate of the local rather than global $N_e$ must account, when possible, for these parameters (Palsbøll et al. 2013). For example, Alter et al. (2007) have not found population subdivision in the eastern Northern Pacific gray whales, but their studies indicate that estimate of $\theta$ in the local population may include the entire Pacific metapopulation. Similarly, Ruegg et al. (2013) using multi-locus sequence data, have not found population subdivision in the Northern Atlantic humpbacks.

Worldwide population structure analysis based on mtDNA showed genetic differentiation among humpback whales from three oceans, the North Atlantic, the North Pacific and the Southern hemisphere. Also, the humpback whales from the Southern Hemisphere are less strongly differentiated from those in the North Atlantic and most differentiated from those in the North Pacific (Baker et al. 1993), as it was also suggested by actin intron sequences analysis (Palumbi and Baker 1994). This suggests that Brazil and Alaska humpback populations, respectively in the South Atlantic and North Pacific Oceans, are differentiated. However, our Bayesian clustering analysis was not able to detect a strong signal of population structure with these multiple nuclear loci, suggesting that more polymorphic loci and/or more samples need to be added to this kind of analysis. Anyway, we thought important to take into account the migration parameter in our estimates of $\theta$ using

*Lamarc* and *Migrate*. Although the migration rate estimates were largely uninformative (large 95% confidence intervals), Beerli (2006) verified that precise and accurate estimates of $\theta$ are recovered even when there is no enough information in the data to recover significant migration rate estimates. Most importantly, when $\theta$ was estimated with each population separately, our estimates become slightly higher (8% for the BSA humpback whales and tightly overestimated at 24% for the GB humpbacks, data not shown) than those with both populations with migration. This strengthens again the importance of accounting for migration when estimating $\theta$.

Recent studies have reported a weak population differentiation among the humpback whale breeding grounds in the Southern Hemisphere (Olavarría et al. 2007, Rosenbaum et al. 2009). Rosenbaum et al. (2009) have showed low genetic differentiation (BSA - BSB1, $F_{ST}$ = 0.0073; BSA - BSB2, $F_{ST}$ = 0.0098) and high migration rates (26 migrants per generation) between the stocks (BSA and BSB) from the South Atlantic Ocean based on the mtDNA analysis. In addition, more recently Ruegg et al. (2013) have indicated a significant population structure between the humpback whales from North Atlantic and South Atlantic Oceans, and a lack of significant population structure within each ocean basin based on multi-locus analysis. Therefore, it is possible that our estimate of long-term population size for the Brazilian humpback whales may be including at least the whole South Atlantic metapopulation.

We also estimated the divergence time between the two populations using the IMa2 method, which also estimates the ancestral and the two descendent population sizes as well as the gene flow. The results suggest that the two populations, Glacier Bay and BSA representing the humpback whale populations of the North Pacific and South Atlantic Oceans, respectively, have diverged from a common ancestral population at approximately 34,000 years ago, at the end of the Pleistocene period and before the Last Glacial Maximum (LGM), although the confidence interval for this estimate is very broad (~10,000-110,000 years ago). This divergence between North Pacific and South Atlantic humpback populations during the Last Glacial Period (10,000-110,000 years ago) suggests a possible influence of the glacial climate in the evolutionary history of these populations. However, our point estimate around 40 kya for this divergence, although preliminary, suggests that the divergence between these populations may be much more recent than previously thought (e.g. Alter et al. 2007, Ruegg et al. 2013). Despite the existence of occasional trans-equatorial and trans-oceanic migrations among humpback whale populations, this relatively recent population separation may explain the low level of differentiation between Southern and Northern Hemispheres stocks as

detected here and elsewhere for other data sets (Baker et al. 1993, 1998, Medrano-González et al. 2001, Olavarría et al. 2007, Palsbøll et al. 1995, Rosenbaum et al. 2009).

Finally, the historical population size dynamics estimated using the extended Bayesian skyline plot has recuperated a very interesting signal of population decline over the last ~30 kya and a signal of a very recent (~200 years) population decline, the latter coincident with the onset of the anthropogenic exploitation in the region (Fig. 2). As many authors have emphasized the role of multiple loci in recovering population dynamics (Carling and Brumfield 2007, Heled and Drummond 2007), this result suggests that the large number of loci (35 nuclear loci and mtDNA control region) used in this study increased the statistical power to detecting a very recent population decline. The only other study that have somewhat detected signal of a recent genetic bottleneck for this population was our previous study using different methods (Cypriano-Souza et al. in prep.). The effective population size before the whaling reduction was estimated in about 30,000 individuals, which corresponds to the distribution of 40,886 (27,777-64,031) estimated by *Lamarc*, although the confidence interval of the former is large.

Concerning the population dynamics during the Last Glacial Period (110,000 to 10,000 years ago), it is notable that the expansion began before the onset of this period, and the subsequent population decline was estimated to have started at the onset of the LGM (~30,000). This may be explained by strong influence of the glacial climate changes, sea level oscillations and consequently the ocean productivity affecting the whale populations. Behrenfeld et al. (2006) have indicated that ocean productivity is largely determined by temperature which are reduced in a warmer period. Therefore, this suggests probably that the relationship between whale population size and climate change is mainly related with ecosystem carrying capacity. More recently, Phillips et al. (2012) have found a similar population dynamics for the bowhead whales that began the expansion about 75,000 years ago and followed by a subsequent bottleneck at the end of LGM (~15,000 years ago).

Overall, the results presented in this study have complemented our previous study (Cypriano-Souza et al. in prep) to improving the understanding of the demographic history of the humpback whale population wintering off Brazilian coast. With these multiple loci data for humpback whale we have been able, for the first time, to estimate the demographic history of this population during the end of the Pleistocene. Interestingly these new results are the first evidence that Southwestern Atlantic (more likely the whole South Atlantic metapopulation) was declining at the onset of the whaling, which may explain the discrepancies between the catch records and genetic estimates of the census size at that

period. Very significantly, we have found for the first time, a signal for a very recent population decline, coincident with anthropogenic population depletion. Our genetic-based estimate of historical abundance indicates that the Southwestern Atlantic humpback whale population was much larger than previously estimated by whaling catch records, and corroborating the high long-term census size estimated for the humpback whales from North Atlantic Ocean (Ruegg et al. 2013). Despite evolutionary historical data have been overlooked from most conservation management plans, our findings suggest that this information provides a background for determining the potential impacts of the environmental conditional on an ecologically important and threatened species, mainly with the increasing of the global warming in the last few years. Additionally, future investigations of the effects of humpback whale population structure in the South Atlantic Ocean and reducing some uncertainty factors may improve both genetic and whaling catch-based estimates.

**Funding**

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [to A.L.C.-S.]; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [to S.L.B.]; Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) [to S.L.B.]; Instituto Chico Mendes de Conservação da Biodiversidade/Ministério do Meio Ambiente (ICMBIO/MMA); and Petróleo Brasileiro S.A (PETROBRAS). Special support was given by CAPES for the scholarship award to A.L.C.-S. to conduct part of her doctoral research as a visiting student at Oregon State University (OSU). Biopsy samples were collected off Brazil under permit SISBIO 21489-1.

**Acknowledgements**

We are grateful to the staff of Instituto Baleia Jubarte for collecting Brazilian humpback whale samples in the field, and to the Structure of Populations, Levels of Abundance, and Status of Humpbacks (SPLASH) for providing DNA samples of the Southeastern Alaska humpback whale. We thank Janice M. Straley and Christine M. Gabriele for their suggestions in the selection of the Alaska humpback whale samples. We also thank David Jacobsen for assistance with 454 sequencing, Alana Alexander, Angie Sremba, Renee Gibb, Rebecca Hamner and Debbie Steel for laboratory assistance.

**References**


Aitken N, Smith S, Schwarz C, Morin PA. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol Ecol. 13:1423-1431.

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

Alter SE, Rynes E, Palumbi SR. 2007. DNA evidence for historic population size and past ecosystem impacts of gray whales. Proc Natl Acad Sci USA. 104:15162-15167.

Alter SE, Palumbi SR. 2009. Comparing evolutionary patterns and variability in the mitochondrial control region and cytochrome b in three species of baleen whales. J Mol Evol. 68:97-111.

Alter SE, Newsome SD, Palumbi SR. 2012. Prewhaling genetic diversity and population ecology in eastern pacific grey whales: insights from ancient DNA and stable isotopes. PLoS ONE 7:e35039.

Andriolo A, Martins CCA, Engel MH, Pizzorno JL, Más-Rosa S, Freitas A, Morete ME, Kinas PG. 2006. The first aerial survey to estimate abundance of humpback whale (*Megaptera novaeangliae*) in the breeding ground off Brazil. J Cetacean Res Manage. 8:307-311.

Andriolo A, Kinas PG, Engel MH, Martins CCA. 2010. Humpback whale population estimates and distribution along the Brazilian breeding ground. Endang Spec Resear. 11: 233-243.

Baker CS, Perry A, Bannister JL, Weinrich MT, Abernethy RB, Calambokidis J, Lien J, Lambertsen RH, Urban-Ramirez J, Vasquez O, Clapham PJ, Alling A, O´Brien SJ, Palumbi SR. 1993. Abundant mitochondrial DNA variation and worldwide population structure in humpback whales. Proc Natl Acad Sci USA. 90:8239-8243.

Baker CS, Flóres-Gonzáles 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, Clapham PJ. 2004. Modeling the past and future of whales and whaling. Trends Ecol Evol. 19:365-371.

Beerli P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic 274 parameters. Bioinformatics. 22:341-345.

Beerli P. 2008. Migrate version 3.0 - a maximum likelihood and Bayesian estimator of gene flow using the coalescent. Available: http://popgen.scs.edu/ migrate.html.

Behrenfeld MJ, O'Malley RT, Siegel DA, McClain CR, Sarmiento JL, Feldman GC, Milligan AJ, Falkowski PG, Letelier RM, Boss ES. 2006. Climate-driven trends in contemporary ocean productivity. Nature. 444:752-754.

Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willersley E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS ONE. 2:e197.

Carling MD, Brumfield RT. 2007. Gene sampling strategies for multi-Locus population estimates of genetic diversity (Θ). PLoS ONE. 2:e160.

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics. 10:195–205.

Chittleborough RG. 1965. Dynamics of two populations of the humpback whale, Megaptera novaeangllae (Borowski). Aust J Mar Freshw Res. 16:33-128.

Clapham PJ, Baker CS. 2002. Modern whaling. In: Perrin WF, Würsig B, Thewissen JGM, editors. Encyclopedia of Marine Mammals, New York: Academic Press; p. 1328-1332.

Clapham P, Palsboll P, Pastene L, Smith T, Walloe L. 2005. Estimating prewhaling abundance. J Cetacean Res Manage. 7(Suppl.): 386–387.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37(Database issue):D141-145.

Cypriano-Souza AL, Fernández GP, Lima-Rosa CAV, Engel MH, Bonatto SL. 2010. Microsatellite genetic characterization of the humpback whale (*Megaptera novaeangliae*) Breeding Ground off Brazil (Breeding Stock A). J Hered. 101(2):189-200.

Descamps S, Campbell MA. 2010. Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. Mol Breeding. 25 (4):553-570.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resources. 4(2):359-361.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Ekblom R, Galindo J. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity. 107:1-15.

Ellis M. 1969. A baleia no Brasil colonial. Edições Melhoramentos, Editora da Universidade de São Paulo, São Paulo. 235 pp.

Engel MH, Fagundes NJR, Rosenbaum HC, Leslie MS, Ott PH, Schmitt R, Secchi E, Dalla Rosa L, Bonatto SL. 2008. Mitochondrial DNA diversity of the Southwestern Atlantic humpback whale (*Megaptera novaeangliae*) breeding area off Brazil, and the potential connections to Antartic feeding areas. Conserv Genet. 9:1253-1262.

Engel MH, Martin AR. 2009. Feeding grounds of the western South Atlantic humpback whale population. Mar. Mamm. Sci. 25:964–969.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14(8):2611-2620.

Findlay KP. 2001. A review of humpback whale catches by modern whaling operations in the Southern Hemisphere. Mem Quensl Mus. 47(2):411-20.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147:915-925.

Glenn TC. 2011. Field guide to next-generation DNA sequencers. Mol Ecol Resources. 11:759-769.

Goecks J, Nekrutenko A, Taylor J. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology. 11:R86.

Gompert Z, Forister ML, Fordyce JA, Nice CC, Williamson RJ, Alex Buerkle C. 2010. Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of Lycaeides butterflies. Mol Ecol. 19(12):2455-2473.

Harismendy O, Pauline CN, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biology. 10:R32.

Headland R. 1984. The island of South Georgia. London: Cambridge University Press; p 291.

Heled J, Drummond A. 2007. Bayesian inference of population size history from multiple loci. BMC Evol Biol. 8:289.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics. 167:747-760.

Hey J, Won Y-J, Sivasundar A, Nielsen R, Markert JA. 2004. Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. Mol Ecol. 13:909-919.

Hey J. 2010. Isolation with Migration Models for More Than Two Populations. Mol. Biol. Evol. 27:905-920.

Hird SM, Brumfield RT, Carstens BC. 2011. PRGMATIC: an efficient pipeline for collating genome-enriched second-generation sequencing data using a 'provisional reference genome'. Mol Ecol Resources. 11:743-748.

Hird SM. 2012. lociNGS: A Lightweight Alternative for Assessing Suitability of Next-Generation Loci for Evolutionary Analysis. PLoS ONE. 7:10.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Resources. 11(1):117-122.

Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. Genome Research. 9:868-877.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. Mol Ecol Resources. 9:1322–1332.

IUCN2008. 2008 IUCN red list of threatened species. [Internet]. [cited 4 May 2009]. Available from: http://www.iucnredlist.org/.

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.

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-Gonzalez L, Palumbi SR. 2009. Big and slow: phylogenetic estimates of molecular evolution in baleen whales (suborder mysticeti). Mol Biol Evol. 26:2427-2440.

Koboldt D, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Ding L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 25(17):2283-2285.

Kuhner MK. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics. 22:768-770.

Kumar S, Carlsen T, Mevik B-H, Enger P, Blaalid R, Shalchian-Tabrizi K, Kauserud H. 2011. CLOTU: An online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. MC Bioinformatics. 12:182.

Lambertsen RH. 1987. A biopsy system for large whales and its use for cytogenetics. J Mamm. 68:443-445.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 25:1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078-2079.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25:1451-1452.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of Next-Generation Sequencing Systems. J Biomed Biotech. Volume 2012, Article ID 251364.

Lodi L. 1992. Uma história da caça à baleia. Ciência Hoje. 14(81):78-83.

Lubick N. 2003. New count of old whales adds up to big debate. Science 301:5632.

Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, O'Brien SJ. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nature Genetics. 15:47-56.

Margulies M, Egholm M, Altman WE, Attiya S. Bader JS, Bemben LA, Volkmer GA, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437:376–380.

Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. Trends in Genetics. 24(3):133-141.

Martins CCA, Morete ME, Engel MH, Freitas A, Secchi ER, Kinas PG. 2001. Aspects of habitat use patterns of humpback whales in the Abrolhos bank, Brazil, breeding ground. Mem Queensl Mus. 47:563-570.

McCormack JE, Maley JM, Hird SM, Derryberry EP, Graves GR, Brumfield RT. 2012. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Mol Phyl Evol. 62(1):397-406.

Medrano-González L, Baker CS, Robles-Saavedra MR, Murrell J, Vazquez-Cuevas MJ, Congdon BC, Stranley JM, Calambokidis J, Urban-Ramirez J, Flórez-González L, Olavarría C, Aguayo-Lobo A, Nolasco-Soto J, Juarez-Salas RA, Villavicencio-Llamosas K. 2001. Trans-oceanic population genetics structure of humpback whales in the North and South Pacific. Mem Queensl Mus. 47:465-479.

Mesnick SL, Taylor BL, Archer FI, Martien KK, Trevino SE, Hancock-Hanser BL, Medina SCM, Pease VL, Robertson KM, Straley JM, Baird RW, Calambokidis J, Schorr GS, Wade P, Burkanov V, Lunsford CR, Rendell L, Morin PA. 2011. Sperm whale population structure in the eastern and central North Pacific inferred by the use of

single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. Mol Ecol Resources. 11(1):278-298.

Morin PA, Luikart G, Wayne RK. 2004. SNPs in ecology, evolution and conservation. Trends Ecol Evol. 19:208-216.

Morin PA, Aitken NC, Rubio-Cisneros N, Dizon AE, Mesnick S. 2007. Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*). Mol Ecol Notes 7:626-630.

Morin PA, McCarthy M. 2007. Highly accurate SNP genotyping from historical and low-quality samples. Mol Ecol Notes. 7:937-946.

Morin PA, Pease VL, Hancock BL, Robertson KM, Antolik CW. 2010. Characterization of 42 single nucleotide polymorphism (SNP) markers for the bowhead whale (*Balaena mysticetus*) forusein discriminating populations. Mar Mammal Sci. 26: 716-732.

Morin PA, Archer FI, Pease VL, Hancock-Hanser BL, Robertson KM, Huebinger RM, Martien KK, Bickham JW, George JC, Postma LD, Taylor BL. 2012. An empirical comparison of SNPs and microsatellites for population structure, assignment, and demographic analyses of bowhead whale populations. Endang Spec Resear. 19:129-147.

Morozova O, Marra MA. 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics. 92:255-264.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation. A Markov chain Monte Carlo approach. Genetics. 158:885-896.

Niimi S, Watanabe MX, Kim EY, Iwata H, Yasunaga G, Fujise Y, Tanabe S. 2005. Molecular cloning and mRNA expression of cytochrome P4501A1 and 1A2 in the liver of common minke whales (*Balaenoptera acutorostrata*). Mar Pollut Bull. 51:8-12.

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 humpback whales throughout the South Pacific and the origin of the eastern Polynesian breeding grounds. Mar Ecol Prog Ser. 330:257-268.

Palumbi SR, Baker CS. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Bio Evol. 11:426-435.

Palsbøll PJ, Clapham PJ, Mattila DK, Larsen F, Sears R, Siegismund HR, Sigurjónsson J, Vásquez O, Artander P. 1995. Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. Mar Ecol Prog Ser. 116:1-10.

Palsbøll PJ, Peery MZ, Olsen MT, Beissinger SR, Bérubé M. 2013. Inferring recent historic abundance from current genetic diversity. Mol Ecol. 22:22-40.

Phillips CD, Hoffman JI, George JC, Suydam RS, Huebinger RM, Bickham JW, Patton JC. 2012. Molecular insights into the historic demography of bowhead whales: understanding the evolutionary basis of contemporary management practices. Ecol Evol. 3(1):18-37.

Polanowski AM, Schmitt NT, Double MC, Gales NJ, Jarman SN. 2011. TaqMan assays for genotyping 45 single nucleotide polymorphisms in the humpback whale nuclear genome. Conserv Genet Resources. 3(4):645-649.

Pritchard JK, Stephens M, Donnely P. 2000. Inference of a population structure using multilocus genotype data. Genetics. 155:945-959.

Pritchard JK, Wen X, Falush D. 2007. Documentation for structure software: version 2.2. Available from http://pritch.bsd.uchicago.edu/software

Prychitko TM, Moore WS. 2003. Alignment and phylogenetic analysis of betafibrinogen intron 7 sequences among avian orders reveal conserved regions within the intron. Mol Biol Evol. 20:762-771.

Punt AE, Friday NA, Smith TD. 2006. Reconciling data on the trends and abundance of North Atlantic humpback whales within a population modelling framework. J Cetacean Res Manage. 8:145-159.

Puritz JB, Addison JA, Toonen RJ. 2012. Next-Generation phylogeography: A targeted pproach for multilocus sequencing of non-model organisms. PLoS ONE. 7(3):e34241.

Rambaut A, Drummond AJ. 2007. Tracer v1.4. Available: http://beast.bio.ed. ac.uk/Tracer.

Rice DW. 1978. The humpback whale in the North Pacific: distribution, exploitation, and numbers. In: Norris KS, Reeves R. editors. Report on a workshop on problems related to humpback whales (*Megaptera novaeangliae*) in Hawaii. Washington: U.S. Marine Mammal Commission, p. 29-44.

Roman J, Palumbi S. 2003. Whales before whaling in North Atlantic. Science. 301:508-510.

Rooney AP, Honeycutt RL, Davis SK, Derr JN. 1999. Evaluating a putative bottleneck in a population of bowhead whales from patterns of microsatellite diversity and genetic disequilibria. J Mol Evol. 49:682-690.

Rooney AP, Honeycutt RL, Derr JN. 2001. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. Evolution. 55:1678-1685.

Rosenbaum HC, Pomilla C, Mendez M, Leslie MS, Best PB, Findlay KP, Minton G, Ersts PJ, Collins T, Engel MH, Bonatto SL, Kotze DPGH, Meÿer M, Barendse J, Thornton M,

Razafindrakoto Y, Ngouessono S, Vely M, Kiszka J. 2009. Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. PLoS ONE. 4:e7318.

Ruegg KC, Anderson EC, Baker CS, Vant M, Jackson JA, Palumbi SR. 2010. Are Antarctic minke whales unusually abundant because of 20th century whaling? Mol Ecol. 19:281-291.

Ruegg KC, Rosenbaum HC, Anderson EC, Engel MH, Rothchild A, Baker CS, Palumbi SR. 2013. Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure. Conserv Genet. 14:103-114.

Stevick PT, Pacheco de Godoy L, McOsker M, Engel MH, Allen A. 2006. A note on the movement of a humpback whale from Abrolhos Bank, Brazil to South Georgia. J Cetacean Res Manage. 8:297–300.

Stuglik MT, Radwan J. Babik W. 2011. jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. Mol Ecol Resources. 11:739–742.

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

Takahata N. 1986. An attempt to estimate the effective size of the ancestral species common to two extant species from which homologous genes are sequenced. Genetical Research, Cambridge. 48:187-190.

Takahata N, Satta Y, Klein J. 1995. Divergence time and population size in the lineage leading to modern humans. Theoretical Population Biology. 48:198-221.

Taylor BL, Chivers SJ, Larese J, Perrin WF. 2007. Generation length and percent mature estimates for IUCN assessments of cetaceans. National Marine Fisheries Service, Southwest Fisheries Science Center.

Tønnessen JN, Johnsen AO. 1982. The history of modern whaling. Berkeley (CA): University of California Press.

Wade PR, Perryman W. 2002. An assessment of the eastern gray whale population in 2002. Unpublished paper SC/54/BRG7 presented to the Scientific Committee of the International Whaling Commission.

Waldick RC, Kraus S, Brown M, White BN. 2002. Evaluating the effects of historic bottleneck events: an assessment of microsatellite variability in the endagered, North Atlantic right whale. Mol Ecol. 11:2241-2249.

Waples RS. 2010. Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. Mol Ecol Resource. 10:785-796.

Ward E, Zerbini, AN, Kinas PG, Engel MH, Andriolo A. 2011. Estimates of population growth rates of humpback whales (*Megaptera novaeangliae*) in the wintering grounds off the coast of Brazil (Breeding Stock A). J Cet Res Manage. 3:145-149.

Wedekin LL, Engel MH, Azevedo A, Kinas PG, Andriolo A, Simões-Lopes PC. 2010. Density and abundance of the humpback whale in the Brazilian breeding ground (stock A): aerial survey, 2008. Unpublished paper SC/62/SH28 presented to the Scientific Committee of the International Whaling Commission.

Yablokov AV. 1994. Validity of whaling data. Nature. 367:108.

Yablokov AV, Zemsky VA, Mikhalev YA, Tornosov VV, Berzin AA. 1998. Data on Soviet whaling in the Antarctic in 1947-1972 (population aspects). Russ J Ecol. 29:38-42.

Yang ZH. 1997. On the estimation of ancestral population sizes of modern humans. Genet Res. 69:111-116.

Zellmer AJ, Hanes MM, Hird SM, Carstens BC. 2012. Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. Syst Biol. 6(5):763-777.

Zerbini AN, Andriolo A, da Rocha JM, Simões-Lopes PC, Siciliano S, Pizzorno JL, Waite JM, DeMaster DP, VanBlaricon GR. 2004. Winter distribution and abundance of humpback whales (*Megaptera novaeangliae*) off Northeastern Brazil. J. Cetacean Res. Manage. 6:101-107.

Zerbini AN, Andriolo A, Heide-Jorgensen MP, Pizzorno JL, Maia YG, VanBlaricom GR, DeMaster DP, Simões-Lopes PC, Moreira S, Bethlem C. 2006a. Satellite-monitored movements of humpback whales *Megaptera novaeangliae* in the Southwest Atlantic Ocean. Mar Ecol Prog Ser. 313:295-304.

Zerbini AN, Ward E, Engel MH, Andriolo A, Kinas PG. 2006b. A Bayesian Assessment of the Conservation Status of Humpback Whales (*Megaptera novaeangliae*) in the Western South Atlantic Ocean (Breeding Stock A). Unpublished paper SC/58/SH2 presented to the Scientific Committee of the International Whaling Commission.

Zerbini AN, Andriolo A, Danilewicz D, Heide-Jorgensen MP, Gales N, Clapham PJ. 2011. An update on research on migratory routes and feeding destinations of Southwest Atlantic humpback whales. Unpublished paper SC/63/SH23 presented to the Scientific Committee of the International Whaling Commission.

**Figure Legends**

Figure 1. STRUCTURE barplot of the proportional membership (q) of each individual of humpback whale in each cluster for $K = 2$ (a) and $K = 3$ (b) without use of location prior, and for $K = 2$ (c) and $K = 3$ (d) with sampling location prior. Each individual is represented by a vertical bar broken into colored segments with the length indicating the coefficient of membership to each population. Black line represents the boundary between individuals sampled in each of the two regions (Brazil and Alaska).

Figure 2. Extend Bayesian skyline plots depicting the effective population size fluctuations over time at different time frames. (a) Timing past 500,000 years, highlighting an expansion followed by a population decline, (b) timing past 100,000 years, showing a population decline beginning around 30,000 years ago, and (c) time past 5,000 years, emphasizing a bottleneck approximately 200 years ago (~11 generations ago). Blue solid line represents the median estimates, and the yellow area denotes the 95% highest posterior densities (HPDs) for the estimates.

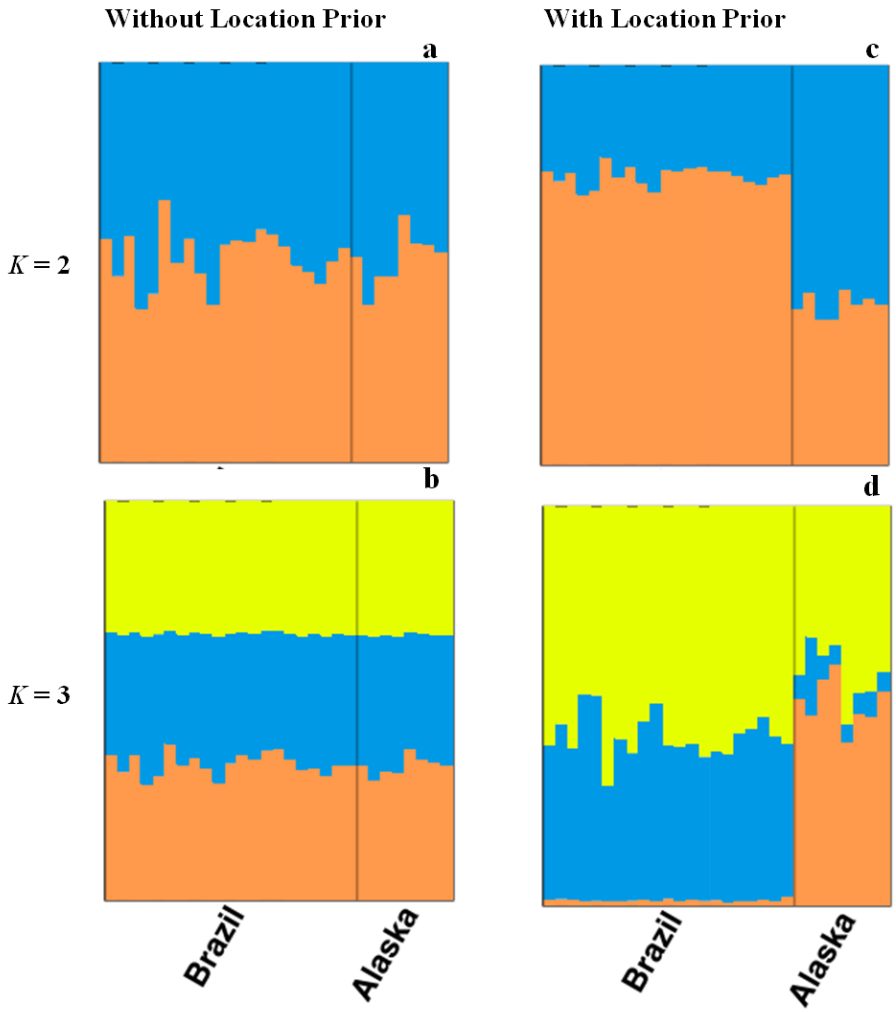
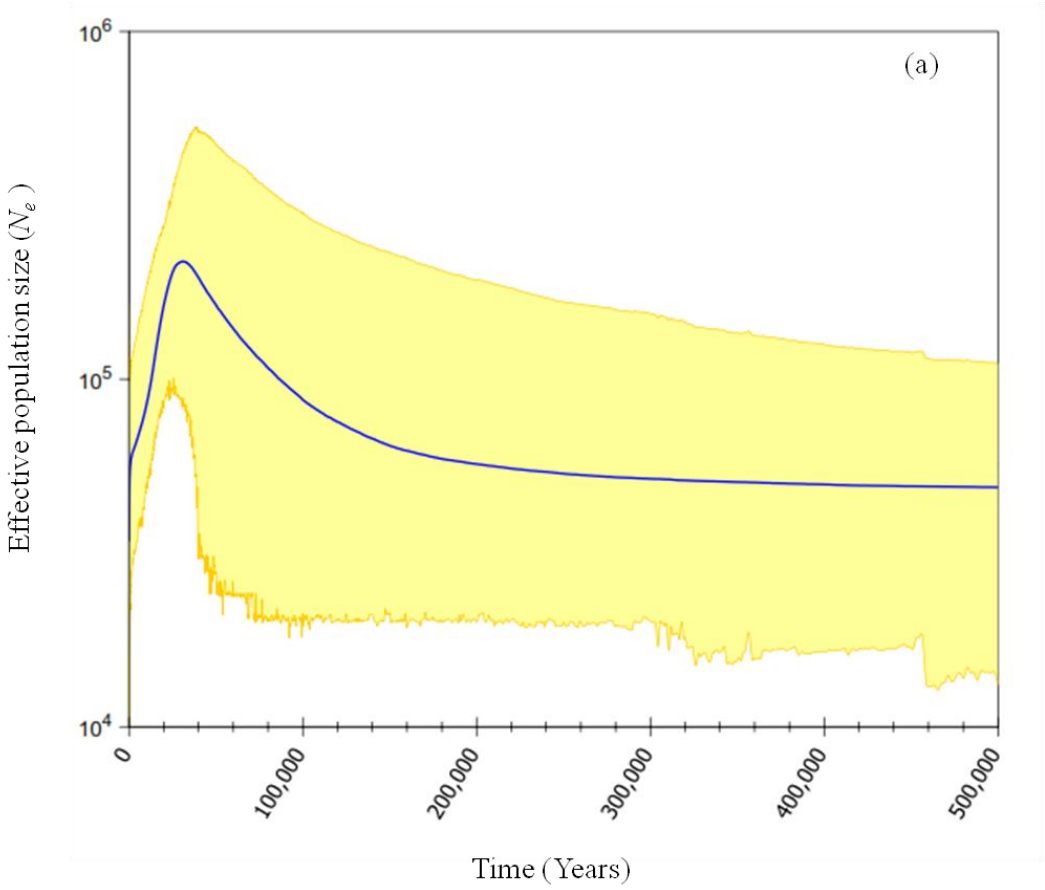

Without Location Prior
With Location Prior
a
c
K = 2
b
d
K = 3
Brazil
Alaska
Brazil
Alaska

Figure 1


(a)
Effective population size ($N_e$)
$10^6$
$10^5$
$10^4$
0
100,000
200,000
300,000
400,000
500,000
Time (Years)

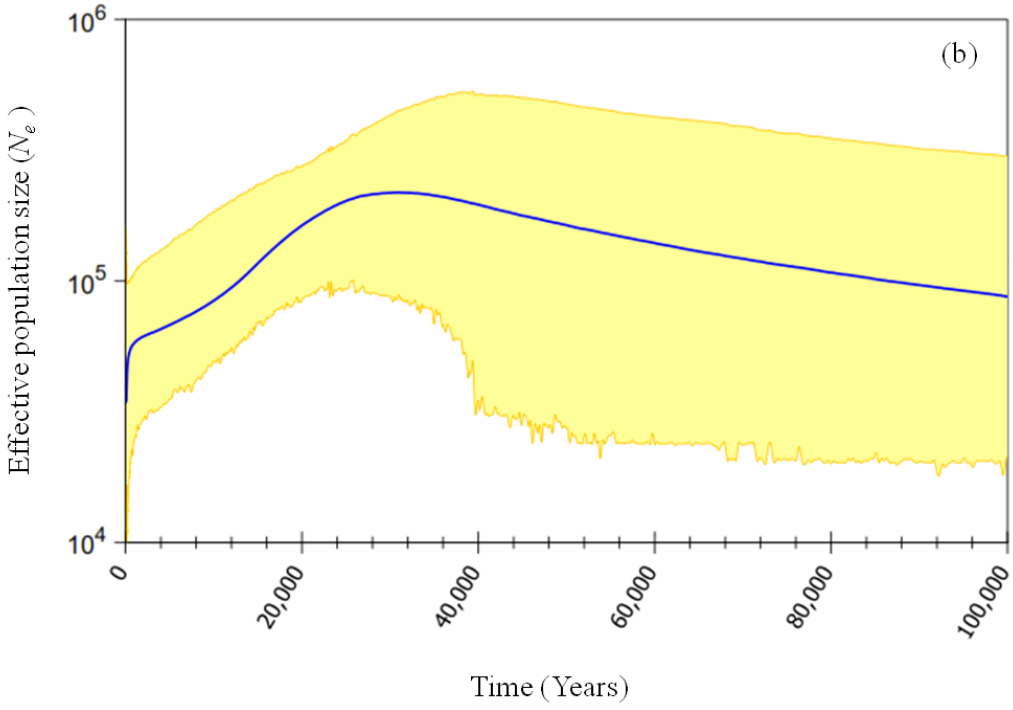

(b)
Effective population size ($N_e$)
$10^6$
$10^5$
$10^4$
0
20,000
40,000
60,000
80,000
100,000
Time (Years)

Figure 2

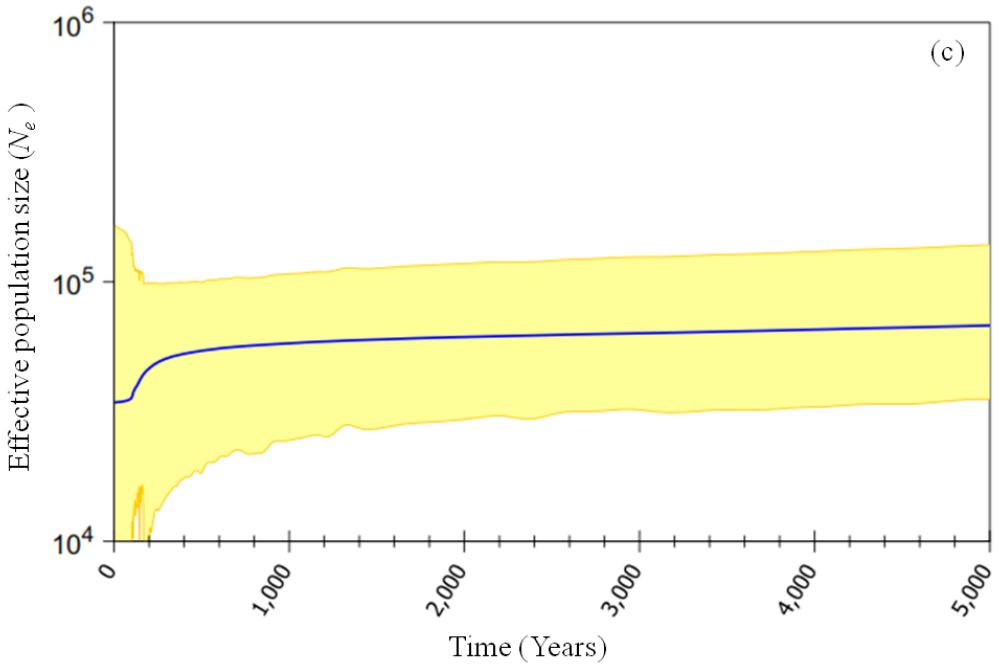



Figure 2 (Continued)

Table 1. Results from 454 sequencing, diagnostic pipeline (RDP), and loci validation.

| | LB1 | LB2 | Total |
|---|---|---|---|
| Total reads | 50,480 | 80,232 | 130,712 |
| Total QC reads > 150 bp (% of total) | 40,171 (79.6%) | 56,507 (70.4%) | 96,678 (73.9%) |
| Total QC reads with targeted loci (% of total QC) | 36,791 (91.5%) | 50,862 (90.0%) | 87,653 (90.6%) |
| Total loci | 22 | 22 | 44 |
| Total QC reads validated | 36,374 | 50,016 | 86,390 |
| Total loci validated | 21 | 18 | 39 |
| Total length | 9,087 bp | 8,055 bp | 17,142 bp |
| Average read length | 432 bp | 447 bp | 440 bp |

LB1 and LB2 - libraries 1 and 2, respectively.

Table 2. List of 39 loci sequenced and validated, showing the sequence length, number of individuals (gene copies), number of variable sites, indels, and alleles for each locus.

| Locus | Amplicon length (bp) | No. Ind. (No. Gene copies) | Variable sites | Indels (bp) | Alleles |
|---|---|---|---|---|---|
| BH1 | 461 | 29 (58) | 4 | 0 | 4 |
| BH4 | 381 | 22 (44) | 3 | 0 | 3 |
| BH31 | 477 | 29 (58) | 0 | 0 | 1 |
| BH42a | 407 | 26 (52) | 3 | 0 | 3 |
| BH42b | 424 | 26 (52) | 5 | 1 (1) | 7 |
| BH43 | 579 | 28 (56) | 3 | 0 | 4 |
| BH60 | 430 | 29 (58) | 4 | 0 | 5 |
| BH96 | 473 | 23 (46) | 3 | 0 | 4 |
| BH108 | 405 | 28 (56) | 0 | 0 | 1 |
| BH368 | 373 | 23 (46) | 1 | 0 | 2 |
| BH382 | 482 | 29 (58) | 2 | 0 | 2 |
| BH395 | 399 | 26 (52) | 8 | 0 | 4 |
| BH404 | 413 | 24 (48) | 3 | 0 | 4 |
| BH412 | 503 | 27 (54) | 0 | 0 | 1 |
| ACT | 511 | 29 (58) | 7 | 0 | 4 |
| CHRNA1 | 386 | 29 (58) | 3 | 0 | 4 |
| CYP1A1 | 509 | 25 (50) | 1 | 0 | 2 |
| FIN275 | 352 | 29 (58) | 1 | 0 | 2 |
| FIN983 | 387 | 26 (52) | 3 | 0 | 4 |
| FIN3574 | 368 | 29 (58) | 1 | 1 (2) | 3 |
| FIN4225 | 367 | 20 (40) | 1 | 0 | 2 |
| FIN 401 | 431 | 18 (36) | 3 | 0 | 3 |
| FIN1264 | 450 | 27 (54) | 6 | 1 (5) | 8 |
| FIN1718 | 412 | 30 (60) | 2 | 0 | 3 |
| FIN1890 | 481 | 28 (56) | 2 | 0 | 3 |
| FIN2032 | 457 | 25 (50) | 1 | 0 | 2 |
| FIN2057 | 440 | 30 (60) | 3 | 0 | 4 |
| FIN2180 | 419 | 30 (60) | 1 | 0 | 2 |
| FIN3028 | 446 | 29 (58) | 2 | 1 (4) | 4 |
| FIN3108 | 426 | 28 (56) | 3 | 1 (1) | 4 |
| FIN3875 | 431 | 28 (56) | 0 | 0 | 1 |
| FIN3971 | 483 | 29 (58) | 1 | 0 | 2 |
| FIN4082 | 457 | 30 (60) | 4 | 1 (2) | 6 |
| FIN4442 | 497 | 28 (56) | 5 | 0 | 8 |
| FIN4591 | 437 | 27 (54) | 0 | 1 (5) | 2 |
| FIN5488 | 495 | 26 (52) | 4 | 0 | 7 |
| FIN6817 | 447 | 28 (56) | 1 | 1 (2) | 2 |
| FIN6845 | 417 | 28 (56) | 3 | 0 | 3 |
| FIN6887 | 429 | 27 (54) | 2 | 0 | 3 |
| Total | 17,142 | - | 99 | 8 | 133 |

Table 3. Summary statistics for 39 loci sequenced for the Southwestern Atlantic (BSA) and Southeastern Alaska (GB) humpback whales. Includes number of samples (gene copies), number of variable sites, indels, and alleles for each locus.

| Locus | Region | N (Gene copies) | Variable sites | Indels (bp) | Alleles | Rm | Π | Tajima's D | Fu's FS |
|---|---|---|---|---|---|---|---|---|---|
| BH1 | BSA | 21 (42) | 4 | 0 | 4 | 0 | 0.00136 | -0.773 | -0.438 |
| | GB | 8 (16) | 3 | 0 | 3 | 0 | 0.00224 | 0.414 | 1.068 |
| BH4 | BSA | 15 (30) | 3 | 0 | 3 | 0 | 0.00052 | -1.731 | -1.627 |
| | GB | 7 (14) | 0 | 0 | 1 | 0 | 0 | - | - |
| BH31 | BSA | 21 (42) | 0 | 0 | 1 | 0 | 0 | - | - |
| | GB | 8 (16) | 0 | 0 | 1 | 0 | 0 | - | - |
| BH42a | BSA | 19 (38) | 3 | 0 | 3 | 0 | 0.00082 | -0.822 | -0.008 |
| | GB | 7 (14) | 0 | 0 | 1 | 0 | 0 | - | - |
| BH42b | BSA | 18 (36) | 4 | 0 | 5 | 0 | 0.00089 | -1.494 | -3.137 |
| | GB | 8 (16) | 4 | 1 (1) | 5 | 0 | 0.00327 | 0.461 | 0.483 |
| BH43 | BSA | 20 (40) | 3 | 0 | 4 | 0 | 0.00130 | -0.333 | -0.579 |
| | GB | 8 (16) | 2 | 0 | 3 | 0 | 0.00112 | 0.200 | 0.112 |
| BH60 | BSA | 21 (42) | 3 | 0 | 3 | 0 | 0.00169 | 0.085 | 1.092 |
| | GB | 8 (16) | 3 | 0 | 3 | 0 | 0.00087 | -1.696 | -0.898 |
| BH96 | BSA | 16 (32) | 2 | 0 | 3 | 0 | 0.00131 | 0.495 | 0.543 |
| | GB | 7 (14) | 3 | 0 | 4 | 0 | 0.00188 | -0.172 | -0.674 |
| BH108 | BSA | 20 (40) | 0 | 0 | 1 | 0 | 0 | - | - |
| | GB | 8 (16) | 0 | 0 | 1 | 0 | 0 | - | - |
| BH368 | BSA | 16 (32) | 1 | 0 | 2 | 0 | 0.00061 | -0.138 | 0.331 |
| | GB | 7 (14) | 0 | 0 | 1 | 0 | 0 | - | - |
| BH382 | BSA | 21 (42) | 2 | 0 | 2 | 0 | 0.00020 | -1.482 | -0.701 |
| | GB | 8 (16) | 0 | 0 | 1 | 0 | 0 | - | - |
| BH395 | BSA | 18 (36) | 8 | 0 | 4 | 0 | 0.00728 | 1.477 | 4.436 |
| | GB | 8 (16) | 6 | 0 | 2 | 0 | 0.00602 | 1.120 | 5.659 |
| BH404 | BSA | 16 (32) | 2 | 0 | 3 | 0 | 0.00123 | 0.038 | 0.128 |
| | GB | 8 (16) | 3 | 0 | 3 | 0 | 0.00349 | 1.723 | 1.874 |
| BH412 | BSA | 19 (38) | 0 | 0 | 1 | 0 | 0 | - | - |
| | GB | 8 (16) | 0 | 0 | 1 | 0 | 0 | - | - |
| ACT | BSA | 21 (42) | 7 | 0 | 4 | 0 | 0.00537 | 1.882 | 4.486 |
| | GB | 8 (16) | 6 | 0 | 3 | 0 | 0.00592 | **2.303** | 4.332 |
| CHRNA1 | BSA | 21 (42) | 3 | 0 | 4 | 0 | 0.00202 | 0.260 | 0.095 |
| | GB | 8 (16) | 0 | 0 | 1 | 0 | 0 | - | - |
| CYP1A1 | BSA | 20 (40) | 1 | 0 | 2 | 0 | 0.00010 | -1.124 | -1.450 |
| | GB | 5 (10) | 0 | 0 | 1 | 0 | 0 | - | - |
| FIN275 | BSA | 21 (42) | 0 | 0 | 1 | 0 | 0 | - | - |
| | GB | 8 (16) | 1 | 0 | 2 | 0 | 0.00092 | 0.155 | 0.551 |
| FIN983 | BSA | 18 (36) | 2 | 0 | 3 | 0 | 0.00042 | -1.284 | -1.893 |
| | GB | 8 (16) | 1 | 0 | 2 | 0 | 0.00060 | -0.448 | 0.083 |

Table 3. Continued.

| Locus | Region | N (Gene copies) | Variable sites | Indels (bp) | Alleles | Rm | Π | Tajima's D | Fu's FS |
|---|---|---|---|---|---|---|---|---|---|
| FIN3574 | BSA | 21 (42) | 1 | 1 (2) | 3 | 0 | 0.00013 | -1.119 | -1.491 |
| | GB | 8 (16) | 0 | 1 (2) | 2 | 0 | 0 | - | - |
| FIN4225 | BSA | 12 (24) | 1 | 0 | 2 | 0 | 0.00023 | -1.159 | -1.028 |
| | GB | 8 (16) | 1 | 0 | 2 | 0 | 0.00089 | 0.155 | 0.551 |
| FIN 401 | BSA | 13 (26) | 3 | 0 | 3 | 0 | 0.00160 | -0.302 | 0.620 |
| | GB | 5 (10) | 2 | 0 | 2 | 0 | 0.00165 | 0.018 | 1.523 |
| FIN1264 | BSA | 20 (40) | 6 | 1 (5) | 8 | 1 | 0.00438 | 1.018 | -0.864 |
| | GB | 7 (14) | 4 | 0 | 4 | 0 | 0.00239 | -0.473 | -0.259 |
| FIN1718 | BSA | 22 (44) | 2 | 0 | 3 | 0 | 0.00033 | -1.304 | -2.149 |
| | GB | 8 (16) | 0 | 0 | 1 | 0 | 0 | - | - |
| FIN1890 | BSA | 22 (44) | 2 | 0 | 3 | 0 | 0.00094 | -0.033 | 0.092 |
| | GB | 6 (12) | 1 | 0 | 2 | 0 | 0.00063 | -0.194 | 0.297 |
| FIN2032 | BSA | 18 (36) | 1 | 0 | 2 | 0 | 0.00024 | -0.813 | -0.597 |
| | GB | 7 (14) | 0 | 0 | 1 | 0 | 0 | - | - |
| FIN2057 | BSA | 22 (44) | 2 | 0 | 3 | 0 | 0.00075 | -0.520 | -0.515 |
| | GB | 8 (16) | 1 | 0 | 2 | 0 | 0.00028 | -1.162 | -0.700 |
| FIN2180 | BSA | 22 (44) | 1 | 0 | 2 | 0 | 0.00118 | 1.601 | 1.843 |
| | GB | 8 (16) | 1 | 0 | 2 | 0 | 0.00119 | 1.308 | 1.247 |
| FIN3028 | BSA | 21 (42) | 2 | 1 (4) | 4 | 0 | 0.00042 | -1.128 | -1.608 |
| | GB | 8 (16) | 1 | 1 (4) | 3 | 0 | 0.00028 | -1.162 | -0.700 |
| FIN3108 | BSA | 21 (42) | 3 | 1 (1) | 4 | 0 | 0.00130 | -0.453 | -0.732 |
| | GB | 7 (14) | 1 | 0 | 2 | 0 | 0.00124 | 1.434 | 1.251 |
| FIN3875 | BSA | 20 (40) | 0 | 0 | 1 | 0 | 0 | - | - |
| | GB | 8 (16) | 0 | 0 | 1 | 0 | 0 | - | - |
| FIN3971 | BSA | 21 (42) | 1 | 0 | 2 | 0 | 0.00037 | -0.338 | 0.085 |
| | GB | 8 (16) | 1 | 0 | 2 | 0 | 0.00067 | 0.155 | 0.551 |
| FIN4082 | BSA | 22 (44) | 4 | 0 | 5 | 0 | 0.00142 | -0.692 | -1.408 |
| | GB | 8 (16) | 0 | 1 (2) | 2 | 0 | 0 | - | - |
| FIN4442 | BSA | 20 (40) | 5 | 0 | 7 | 1 | 0.00192 | -0.479 | -2.456 |
| | GB | 8 (16) | 4 | 0 | 7 | 1 | 0.00314 | 0.915 | -2.512 |
| FIN4591 | BSA | 20 (40) | 0 | 1 (5) | 2 | 0 | 0 | - | - |
| | GB | 7 (14) | 0 | 0 | 1 | 0 | 0 | - | - |
| FIN5488 | BSA | 20 (40) | 4 | 0 | 7 | 2 | 0.00179 | -0.131 | -2.714 |
| | GB | 6 (12) | 2 | 0 | 2 | 0 | 0.00122 | -0.248 | 1.384 |
| FIN6817 | BSA | 20 (40) | 1 | 1 (2) | 2 | 0 | 0.00059 | 0.160 | 0.640 |
| | GB | 8 (16) | 1 | 1 (2) | 2 | 0 | 0.00028 | -1.162 | -0.700 |
| FIN6845 | BSA | 20 (40) | 3 | 0 | 3 | 0 | 0.00087 | -1.062 | -0.393 |
| | GB | 8 (16) | 0 | 0 | 1 | 0 | 0 | - | - |
| FIN6887 | BSA | 19 (38) | 2 | 0 | 3 | 0 | 0.00092 | -0.329 | -0.257 |
| | GB | 8 (16) | 2 | 0 | 3 | 0 | 0.00109 | -0.577 | -0.505 |

*Rm* minimum number of recombination events, $\pi$ nucleotide diversity. Bold numbers refer to a significant deviation from neutrality expectation before a bonferroni correction ($p < 0.05$).

Table 4. Results from *Lamarc* and *Migrate* analyses, showing theta ($\Theta$), effective ($N_e$) and census ($N_c$) population size estimates for each humpback whale population (BSA and GB).

| Method | Population | $\Theta$ (CI) | $N_e$ (CI) | $N_c$ (CI) |
|---|---|---|---|---|
| *Lamarc* | | | | |
| | BSA | 0.0012 (0.0009 - 0.0016) | 40,886 (27,777 - 64,031) | 147,189 (88,886 - 256,124) |
| | GB | 0.0007 (0.0005 - 0.0010) | 24,018 (16,011 - 43,162) | 86,464 (51,235 - 172,648) |
| *Migrate* | | | | |
| | BSA | 0.0013 (0.0012 - 0.0014) | 41,139 (34,188 - 55,335) | 148,100 (109,401 - 221,340) |
| | GB | 0.0005 (0.0004 - 0.0006) | 15,822 (11,396 - 23,715) | 56,959 (36,467 - 94,860) |

CI, 95% confidence or credibility interval.

Table 5. Results from IMa2 analysis, showing parameter estimates (95% highest posterior density) of divergence time (T) between BSA and GB populations, ancestral ($N_{eA}$) and current ($N_{eBSA}$ and $N_{eBSA}$) population sizes, as well as migration rates ($2NM$) between populations (uninformative).

| Parameter | High Point (HPD) | Demographic parameter | High Point (HPD) |
|---|---|---|---|
| $T$ | 0.0066 (0.0018 - 0.0222) | T | 33,883 (9,241 - 113,972) |
| $\Theta_{A}$ | 0.5025 (0.3795 - 0.6435) | $N_{eA}$ | 35,830 (27,060 - 45,884) |
| $\Theta_{BSA}$ | 0.8145 (0.4065 - 2.8690) | $N_{eBSA}$ | 58,007 (28,985 - 204,606) |
| $\Theta_{GB}$ | 0.1065 (0.0135 - 0.3195) | $N_{eGB}$ | 7,594 (962 - 22,782) |
| $m_{BSA>GB}$ | 3.2980 (0.1864 - 3.2980) | $2NM_{BSA>GB}$ | - |
| $m_{GB>BSA}$ | 0.3340 (0.0478 - 3.2980) | $2NM_{GB>BSA}$ | - |

Supplementary Table S1. Primers designed for 47 target loci for the first round PCR.

| Locus | Amplicon length (bp) | F primer sequence | R primer sequence | Reference |
|---|---|---|---|---|
| BH1 | 461 | CCTCTGTCTCTAGATAATGTCCCTG | GAAGAGAGAACTACATCATACCAACAA | Morin et al. 2010 |
| BH4 | 381 | ACGGAACTTATACTATCCTCAACCTAG | GACGTTGTAGAGAATTTACTTCCCT | Morin et al. 2010 |
| BH31 | 477 | AGTGCGTAAGAGTATCTACATTTGC | CCAACAATGGTGTGACCATT | Morin et al. 2010 |
| BH42a | 407 | GAATGTTTTGAAATGAAAGGATAATCC | ACAATCAAGGAGATTATTAAAGCAACATA | Morin et al. 2010 |
| BH42b | 424 | ACTCTGCAGTCCACAGCTCC | CGAGGTTCTACGCCTCGAC | Morin et al. 2010 |
| BH43 | 579 | AGAATTTACTCATAGCCCTGAATAACA | CACATGCAACTGGCTAATATACAC | Morin et al. 2010 |
| BH60 | 430 | TCCTCCAGTGGAACTACTCTCA | CTCAATCCAGGGACAGACAG | Morin et al. 2010 |
| BH96 | 473 | GGGAGGAAGCATAAACTAGAAAA | TCTACAAACTTAATGAGAATAAACCTATATAATAC | Morin et al. 2010 |
| BH108 | 405 | CAAGAGTTTGGTTATAAATGACCCA | CTGAATTCTAATGAACTAACTATATTAAAAAGTT | Morin et al. 2010 |
| BH368 | 373 | TAGATGTCAGGGTCGAAGCA | CACTTGTGGATATGAAATTCTGG | Morin et al. 2010 |
| BH382 | 482 | CACATACAAATGTGGTCAGCA | TCACTGTCACTCCGAAGTTTCT | Morin et al. 2010 |
| BH395 | 539 | ACACTGGAGATTTTTATTATCCCTGTA | CTCGAATGACCTGCATCG | Morin et al. 2010 |
| BH404 | 413 | CAGAAAGCTTCCATAACCACCT | ATCTGTCACATGAGTATGACAAGG | Morin et al. 2010 |
| BH412 | 503 | CTTCAGGCAGATGCGGATG | AGGTACGGGGTTATTGCTCA | Morin et al. 2010 |
| BH414 | 415 | GGGGGAACTTGACAGAAATG | GCTTTGTGGAAACATACCAAA | Morin et al. 2010 |
| ACT | 511 | CCACTACTTTAGGCAG | CTTGTGAACTGATTACAGTCC | Palumbi and Baker 1994 |
| CHRNA1 | 386 | GACCATGAAGTCAGACCAGGAG | GGAGTATGTGGTCCATCACCAT | Lyons et al. 1997 |
| CYP1A1 | 509 | GTCCCCAAAGGCCTGAAG | CATATGGCACAGATGACATTGG | Niimi et al. 2005 |
| FIN275 | 352 | TAGCCATCTGCTCTCTAGCC | TCATACGCAGAAGTCAGTCC | Slikas et al. 2013 |
| FIN983 | 387 | TGCTAACCTTACATTTGCCTC | ACCATACCTACATTAACTGCAC | Slikas et al. 2013 |
| FIN3559 | 326 | ATTTCCCCAACTGCCCTTC | AGCCTATCTCTTCTGCGTC | Slikas et al. 2013 |
| FIN3574 | 368 | TTTCTGGCATTATGGCTTCTC | CAACGCAGGTCACCTATTC | Slikas et al. 2013 |
| FIN4225 | 367 | ATTTGCCCAGACACGCAAC | GCTTTGAACCTTCAATGCCC | Slikas et al. 2013 |

Supplementary Table S1. Continued.

| Locus | Amplicon length (bp) | F primer sequence | R primer sequence | Reference |
|---|---|---|---|---|
| FIN 401 | 432 | CCTTAGGTTTGCTATCCCTG | ACATCACACCCTCCAACTCC | This study |
| FIN744 | 490 | GCCTGGAGTACCCTATCATC | CGGTTCAGAGATCAGGAGGT | This study |
| FIN857 | 522 | CCATGAGGAGCAAATGTGTC | AGAGGGGAAGACGTTATCAC | This study |
| FIN1005 | 502 | CAGGATCACCATACCCTCTG | CAATGCCCAAGTTTGTGCTC | This study |
| FIN1101 | 443 | GCATCCTCTCTCGTCAAAGC | ATTTCCCGCTCTGGTTCAGG | This study |
| FIN1233 | 431 | GCAACACCATCTTTCAACGC | GTCTCAATCCCCAAAGCCTG | This study |
| FIN1264 | 451 | TGCCAGATGACCCAGTTACC | TCCCATATCTCCTCCCTCTC | This study |
| FIN1718 | 412 | AAAGAGCCCTTCAATCACCC | CACGGTGAAGCAAGGTCAAT | This study |
| FIN1890 | 482 | TGTGGGCGTTTGTATGTGTG | CTCACCGAGAAGACAGGAAC | This study |
| FIN2032 | 458 | GACACGGCAGTCAGTTTCCT | GTCAGAGAAGGCATCAGGTG | This study |
| FIN2057 | 441 | GCAGTTGTGTCCCTTCAGCA | CTTTTCCCTGAGCCTCTTGG | This study |
| FIN2180 | 420 | GATGAGGTGTCCTTCCACAG | ATTGGTTGCTGACGGGTTGG | This study |
| FIN3028 | 447 | CCCTGTCTCTAATGTTGGCA | TAGCACTTTCAGTCCTTGGC | This study |
| FIN3108 | 427 | CAGCATCAGTCCTTTTCCCG | CCAAGAAGCGTTACAGAGGC | This study |
| FIN3875 | 431 | ACAGAATGGAGATCCCTGGC | GCACTCTGCTTCCCATAGGA | This study |
| FIN3971 | 487 | CAGATGCTCAGTGGGTAGAG | GCTGTGTTATTGAGAGACGC | This study |
| FIN4082 | 458 | GCTTCCCAGTTGTGACCAAG | ACCCCACTCTATTTGCGAGC | This study |
| FIN4442 | 498 | TTGACACTGAGGAATGTGGG | ACGGCTGACGGAGTAATGAG | This study |
| FIN4591 | 438 | TAGCCAACACCTGTCAAGAG | GCACCAACTTACCTTCCCTA | This study |
| FIN5435 | 409 | TCATCATCTCGGTGGTATCC | CTTAGCCTTCTCTCGTGGTG | This study |
| FIN5488 | 496 | AACCCTAACCAACTCCACTG | GGATCACCGTCCAACATCAA | This study |
| FIN6817 | 449 | CTGGCTGATAGGCACTTCTA | ATGCTCACTGAACGTGTGGA | This study |
| FIN6845 | 418 | AGGGAGGAGACTACAAAGTG | AAACCTGGATGCTCAGATGC | This study |
| FIN6887 | 430 | GAATGGCTCAGGTTGTATCC | TTTGTGTGCTCATAGGTCGC | This study |

Supplementary Table S2. 454 fusion primers (32 pairs), including the A and B (forward and reverse) primers, followed by the MID barcodes (in red), and the complement primers of the universal tails (M13F and T7R).

| Sample | Primer order (5' to 3') | Primer sequence 5'-3' |
|---|---|---|
| BR385-06 | UniA_mid01_M13F | CGTATCGCCTCCCTCGCGCCATCAGacgagtgcgtGTAAAACGACGGCCAGT |
| | UniB_mid01_T7R | CTATGCGCCTTGCCAGCCCGCTCAGacgagtgcgtGCTAGTTATTGCTCAGCGG |
| BR413-07 | UniA_mid02_M13F | CGTATCGCCTCCCTCGCGCCATCAGacgctcgacaGTAAAACGACGGCCAGT |
| | UniB_mid02_T7R | CTATGCGCCTTGCCAGCCCGCTCAGacgctcgacaGCTAGTTATTGCTCAGCGG |
| BR556-08 | UniA_mid03_M13F | CGTATCGCCTCCCTCGCGCCATCAGagacgcactcGTAAAACGACGGCCAGT |
| | UniB_mid03_T7R | CTATGCGCCTTGCCAGCCCGCTCAGagacgcactcGCTAGTTATTGCTCAGCGG |
| BR593-09 | UniA_mid04_M13F | CGTATCGCCTCCCTCGCGCCATCAGagcactgtagGTAAAACGACGGCCAGT |
| | UniB_mid04_T7R | CTATGCGCCTTGCCAGCCCGCTCAGagcactgtagGCTAGTTATTGCTCAGCGG |
| BR603-09 | UniA_mid05_M13F | CGTATCGCCTCCCTCGCGCCATCAGatcagacacgGTAAAACGACGGCCAGT |
| | UniB_mid05_T7R | CTATGCGCCTTGCCAGCCCGCTCAGatcagacacgGCTAGTTATTGCTCAGCGG |
| BR606-09 | UniA_mid06_M13F | CGTATCGCCTCCCTCGCGCCATCAGatatcgcgagGTAAAACGACGGCCAGT |
| | UniB_mid06_T7R | CTATGCGCCTTGCCAGCCCGCTCAGatatcgcgagGCTAGTTATTGCTCAGCGG |
| BR618-09 | UniA_mid07_M13F | CGTATCGCCTCCCTCGCGCCATCAGcgtgtctctaGTAAAACGACGGCCAGT |
| | UniB_mid07_T7R | CTATGCGCCTTGCCAGCCCGCTCAGcgtgtctctaGCTAGTTATTGCTCAGCGG |
| BR624-09 | UniA_mid08_M13F | CGTATCGCCTCCCTCGCGCCATCAGctcgcgtgtcGTAAAACGACGGCCAGT |
| | UniB_mid08_T7R | CTATGCGCCTTGCCAGCCCGCTCAGctcgcgtgtcGCTAGTTATTGCTCAGCGG |
| BR637-09 | UniA_mid09_M13F | CGTATCGCCTCCCTCGCGCCATCAGtagtatcagcGTAAAACGACGGCCAGT |
| | UniB_mid09_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtagtatcagcGCTAGTTATTGCTCAGCGG |
| BR639-09 | UniA_mid10_M13F | CGTATCGCCTCCCTCGCGCCATCAGtctctatgcgGTAAAACGACGGCCAGT |
| | UniB_mid10_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtctctatgcgGCTAGTTATTGCTCAGCGG |
| BR643-09 | UniA_mid11_M13F | CGTATCGCCTCCCTCGCGCCATCAGtgatacgtctGTAAAACGACGGCCAGT |
| | UniB_mid11_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtgatacgtctGCTAGTTATTGCTCAGCGG |

Supplementary Table S2. Continued

| Sample | Primer order (5' to 3') | Primer sequence 5'-3' |
|---|---|---|
| BR645-09 | UniA_mid12_M13F | CGTATCGCCTCCCTCGCGCCATCAGtactgagctaGTAAAACGACGGCCAGT |
| | UniB_mid12_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtactgagctaGCTAGTTATTGCTCAGCGG |
| BR663-09 | UniA_mid13_M13F | CGTATCGCCTCCCTCGCGCCATCAGcatagtagtgGTAAAACGACGGCCAGT |
| | UniB_mid13_T7R | CTATGCGCCTTGCCAGCCCGCTCAGcatagtagtgGCTAGTTATTGCTCAGCGG |
| BR677-10 | UniA_mid14_M13F | CGTATCGCCTCCCTCGCGCCATCAGcgagagatacGTAAAACGACGGCCAGT |
| | UniB_mid14_T7R | CTATGCGCCTTGCCAGCCCGCTCAGcgagagatacGCTAGTTATTGCTCAGCGG |
| BR685-10 | UniA_mid15_M13F | CGTATCGCCTCCCTCGCGCCATCAGatacgacgtaGTAAAACGACGGCCAGT |
| | UniB_mid15_T7R | CTATGCGCCTTGCCAGCCCGCTCAGatacgacgtaGCTAGTTATTGCTCAGCGG |
| BR687-10 | UniA_mid16_M13F | CGTATCGCCTCCCTCGCGCCATCAGtcacgtactaGTAAAACGACGGCCAGT |
| | UniB_mid16_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtcacgtactaGCTAGTTATTGCTCAGCGG |
| BR688-10 | UniA_mid17_M13F | CGTATCGCCTCCCTCGCGCCATCAGcgtctagtacGTAAAACGACGGCCAGT |
| | UniB_mid17_T7R | CTATGCGCCTTGCCAGCCCGCTCAGcgtctagtacGCTAGTTATTGCTCAGCGG |
| BR691-10 | UniA_mid18_M13F | CGTATCGCCTCCCTCGCGCCATCAGtctacgtagcGTAAAACGACGGCCAGT |
| | UniB_mid18_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtctacgtagcGCTAGTTATTGCTCAGCGG |
| BR693-10 | UniA_mid19_M13F | CGTATCGCCTCCCTCGCGCCATCAGtgtactactcGTAAAACGACGGCCAGT |
| | UniB_mid19_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtgtactactcGCTAGTTATTGCTCAGCGG |
| BR703-10 | UniA_mid20_M13F | CGTATCGCCTCCCTCGCGCCATCAGacgactacagGTAAAACGACGGCCAGT |
| | UniB_mid20_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtgtactactcGCTAGTTATTGCTCAGCGG |
| BR708-10 | UniA_mid21_M13F | CGTATCGCCTCCCTCGCGCCATCAGcgtagactagGTAAAACGACGGCCAGT |
| | UniB_mid21_T7R | CTATGCGCCTTGCCAGCCCGCTCAGcgtagactagGCTAGTTATTGCTCAGCGG |
| BR711-11 | UniA_mid22_M13F | CGTATCGCCTCCCTCGCGCCATCAGtacgagtatgGTAAAACGACGGCCAGT |
| | UniB_mid22_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtacgagtatgGCTAGTTATTGCTCAGCGG |

Supplementary Table S2. Continued

| Sample | Primer order (5' to 3') | Primer sequence 5'-3' |
|---|---|---|
| BR718-11 | UniA_mid23_M13F | CGTATCGCCTCCCTCGCGCCATCAGtactctcgtgGTAAAACGACGGCCAGT |
| | UniB_mid23_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtactctcgtgGCTAGTTATTGCTCAGCGG |
| BR769-11 | UniA_mid24_M13F | CGTATCGCCTCCCTCGCGCCATCAGtagagacgagGTAAAACGACGGCCAGT |
| | UniB_mid24_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtagagacgagGCTAGTTATTGCTCAGCGG |
| GB02-03 | UniA_mid25_M13F | CGTATCGCCTCCCTCGCGCCATCAGtcgtcgctcgGTAAAACGACGGCCAGT |
| | UniB_mid25_T7R | CTATGCGCCTTGCCAGCCCGCTCAGtcgtcgctcgGCTAGTTATTGCTCAGCGG |
| GB02-05 | UniA_mid26_M13F | CGTATCGCCTCCCTCGCGCCATCAGacatacgcgtGTAAAACGACGGCCAGT |
| | UniB_mid26_T7R | CTATGCGCCTTGCCAGCCCGCTCAGacatacgcgtGCTAGTTATTGCTCAGCGG |
| GB02-07 | UniA_mid27_M13F | CGTATCGCCTCCCTCGCGCCATCAGacgcgagtatGTAAAACGACGGCCAGT |
| | UniB_mid27_T7R | CTATGCGCCTTGCCAGCCCGCTCAGacgcgagtatGCTAGTTATTGCTCAGCGG |
| GB02-08 | UniA_mid28_M13F | CGTATCGCCTCCCTCGCGCCATCAGactactatgtGTAAAACGACGGCCAGT |
| | UniB_mid28_T7R | CTATGCGCCTTGCCAGCCCGCTCAGactactatgtGCTAGTTATTGCTCAGCGG |
| GB02-12 | UniA_mid29_M13F | CGTATCGCCTCCCTCGCGCCATCAGactgtacagtGTAAAACGACGGCCAGT |
| | UniB_mid29_T7R | CTATGCGCCTTGCCAGCCCGCTCAGactgtacagtGCTAGTTATTGCTCAGCGG |
| GB02-17 | UniA_mid30_M13F | CGTATCGCCTCCCTCGCGCCATCAGagactatactGTAAAACGACGGCCAGT |
| | UniB_mid30_T7R | CTATGCGCCTTGCCAGCCCGCTCAGagactatactGCTAGTTATTGCTCAGCGG |
| GB02-19 | UniA_mid31_M13F | CGTATCGCCTCCCTCGCGCCATCAGagcgtcgtctGTAAAACGACGGCCAGT |
| | UniB_mid31_T7R | CTATGCGCCTTGCCAGCCCGCTCAGagcgtcgtctGCTAGTTATTGCTCAGCGG |
| GB02-21 | UniA_mid32_M13F | CGTATCGCCTCCCTCGCGCCATCAGagtacgctatGTAAAACGACGGCCAGT |
| | UniB_mid32_T7R | CTATGCGCCTTGCCAGCCCGCTCAGagtacgctatGCTAGTTATTGCTCAGCGG |

Supplementary Table S3. Simple statistics for each locus, showing the total number of reads, the relative coverage, and the total number of samples.

| LB1 | | | | | LB2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Locus | Amplicon length (bp) | Reads | Relative coverage | Samples | Locus | Amplicon length (bp) | Reads | Relative coverage | Samples |
| BH1 | 461 | 2037 | 1,18 | 29 | FIN 401 | 431 | 349 | 0,13 | 18 |
| BH4 | 381 | 357 | 0,21 | 22 | FIN744* | 489 | 89 | 0,03 | 0 |
| BH31 | 477 | 1983 | 1,14 | 29 | FIN1005* | 501 | 290 | 0,10 | 0 |
| BH42a | 407 | 683 | 0,39 | 26 | FIN1101* | 442 | 62 | 0,02 | 0 |
| BH42b | 424 | 592 | 0,34 | 26 | FIN1233* | 430 | 187 | 0,07 | 0 |
| BH43 | 579 | 2276 | 1,31 | 28 | FIN1264 | 450 | 1167 | 0,42 | 27 |
| BH60 | 430 | 2105 | 1,22 | 29 | FIN1718 | 412 | 1673 | 0,60 | 30 |
| BH96 | 473 | 241 | 0,14 | 23 | FIN1890 | 481 | 1495 | 0,54 | 28 |
| BH108 | 405 | 875 | 0,51 | 28 | FIN2032 | 457 | 2217 | 0,80 | 25 |
| BH368 | 373 | 381 | 0,22 | 23 | FIN2057 | 440 | 2549 | 0,92 | 30 |
| BH382 | 482 | 6113 | 3,53 | 29 | FIN2180 | 419 | 2409 | 0,87 | 30 |
| BH395 | 399 | 2662 | 1,54 | 26 | FIN3028 | 446 | 3015 | 1,09 | 29 |
| BH404 | 413 | 435 | 0,25 | 24 | FIN3108 | 426 | 3949 | 1,42 | 28 |
| BH412 | 503 | 6402 | 3,70 | 27 | FIN3875 | 430 | 1810 | 0,65 | 28 |
| ACT | 511 | 2843 | 1,64 | 29 | FIN3971 | 483 | 3171 | 1,14 | 29 |
| CHRNA1 | 386 | 948 | 0,55 | 29 | FIN4082 | 457 | 2604 | 0,94 | 30 |
| CYP1A1 | 509 | 559 | 0,32 | 25 | FIN4442 | 497 | 914 | 0,33 | 28 |
| FIN275 | 352 | 1037 | 0,60 | 29 | FIN4591 | 437 | 5539 | 1,99 | 27 |
| FIN983 | 387 | 1987 | 1,15 | 26 | FIN5488 | 495 | 3753 | 1,35 | 26 |
| FIN3559* | 326 | 164 | 0,09 | 0 | FIN6817 | 447 | 4123 | 1,48 | 28 |
| FIN3574 | 368 | 1391 | 0,80 | 29 | FIN6845 | 417 | 5589 | 2,01 | 28 |
| FIN4225 | 367 | 467 | 0,27 | 20 | FIN6887 | 429 | 3690 | 1,33 | 27 |
| Total | 9,413 | 36,538 | - | 29 | Total | 9,917 | 50,644 | - | 30 |
| Total validated | 9,087 | 36,374 | - | 29 | Total validated | 8,055 | 50,016 | - | 30 |
| Average | 433 | 1,732 | - | 26 | Average | 447 | 2,779 | - | 27 |

Supplementary Table S4. Theta estimated for each locus per population (BSA and GB) using *Lamarc*.

| Locus | BSA | | | GB | | |
|---|---|---|---|---|---|---|
| | $\theta$ | Min | Max | $\theta$ | Min | Max |
| BH1 | 0.0020 | 0.0005 | 0.0055 | 0.0018 | 0.0003 | 0.0060 |
| BH4 | 0.0021 | 0.0004 | 0.0051 | 0.00001 | 0.00001 | 0.0015 |
| BH42a | 0.0016 | 0.0003 | 0.0050 | 0.00002 | 0.00001 | 0.0013 |
| BH42b | 0.0026 | 0.0006 | 0.0069 | 0.0027 | 0.0007 | 0.0085 |
| BH43 | 0.0013 | 0.0002 | 0.0037 | 0.0010 | 0.0001 | 0.0039 |
| BH60 | 0.0015 | 0.0002 | 0.0044 | 0.0023 | 0.0004 | 0.0073 |
| BH96 | 0.0010 | 0.0001 | 0.0034 | 0.0023 | 0.0004 | 0.0071 |
| BH368 | 0.0006 | 0.00002 | 0.0029 | 0.00001 | 0.00001 | 0.0017 |
| BH382 | 0.0008 | 0.0001 | 0.0030 | 0.00001 | 0.00001 | 0.0012 |
| BH395 | 0.0035 | 0.0013 | 0.0086 | 0.0032 | 0.0009 | 0.0086 |
| BH404 | 0.0011 | 0.0002 | 0.0041 | 0.0014 | 0.0001 | 0.0050 |
| ACT | 0.0025 | 0.0008 | 0.0058 | 0.0026 | 0.0008 | 0.0071 |
| CHRNA1 | 0.0016 | 0.0003 | 0.0051 | 0.00001 | 0.00001 | 0.0013 |
| CYP1A1 | 0.0005 | 0.00002 | 0.0021 | 0.00001 | 0.00001 | 0.0014 |
| FIN275 | 0.00001 | 0.00001 | 0.0012 | 0.0008 | 0.00003 | 0.0040 |
| FIN983 | 0.0014 | 0.0001 | 0.0045 | 0.0007 | 0.00003 | 0.0037 |
| FIN3574 | 0.0006 | 0.00002 | 0.0029 | 0.00001 | 0.00001 | 0.0014 |
| FIN4225 | 0.0007 | 0.00003 | 0.0034 | 0.0007 | 0.00003 | 0.0040 |
| FIN 401 | 0.0016 | 0.0003 | 0.0053 | 0.0014 | 0.0001 | 0.0061 |
| FIN1264 | 0.0041 | 0.0015 | 0.0090 | 0.0029 | 0.0007 | 0.0090 |
| FIN1718 | 0.0011 | 0.0001 | 0.0040 | 0.00002 | 0.00001 | 0.0013 |
| FIN1890 | 0.0008 | 0.0001 | 0.0031 | 0.0006 | 0.00002 | 0.0035 |
| FIN2032 | 0.0005 | 0.00002 | 0.0023 | 0.00001 | 0.00001 | 0.0013 |
| FIN2057 | 0.0011 | 0.0001 | 0.0037 | 0.0006 | 0.00002 | 0.0034 |
| FIN2180 | 0.0004 | 0.00002 | 0.0021 | 0.0007 | 0.00002 | 0.0032 |
| FIN3028 | 0.0012 | 0.0001 | 0.0036 | 0.0008 | 0.00002 | 0.0035 |
| FIN3108 | 0.0018 | 0.0003 | 0.0050 | 0.0006 | 0.00002 | 0.0034 |
| FIN3971 | 0.0004 | 0.00002 | 0.0021 | 0.0006 | 0.00002 | 0.0029 |
| FIN4082 | 0.0021 | 0.0005 | 0.0058 | 0.00003 | 0.00001 | 0.0012 |
| FIN4442 | 0.0034 | 0.0011 | 0.0076 | 0.0044 | 0.0013 | 0.0116 |
| FIN4591 | 0.00001 | 0.00001 | 0.0009 | 0.00001 | 0.00001 | 0.0013 |
| FIN5488 | 0.0035 | 0.0011 | 0.0078 | 0.0012 | 0.0001 | 0.0048 |
| FIN6817 | 0.0005 | 0.00002 | 0.0022 | 0.0006 | 0.00002 | 0.0034 |
| FIN6845 | 0.0017 | 0.0003 | 0.0050 | 0.00003 | 0.00001 | 0.0013 |
| FIN6887 | 0.0012 | 0.0001 | 0.0038 | 0.0014 | 0.0001 | 0.0056 |
| Overall | 0.0012 | 0.0009 | 0.0016 | 0.0007 | 0.0005 | 0.0010 |

## Capítulo 5 - Conclusões gerais

Essa tese abordou principalmente a diferenciação genética e o nível de fluxo gênico entre as populações de baleias jubarte da costa leste e oeste da América do Sul (estoques reprodutivos A e G, respectivamente); e a história demográfica das jubartes do estoque A, no Oceano Atlântico Sul Ocidental. O principal propósito deste estudo foi gerar informações científicas importantes sobre a espécie, que também proporcionarão subsídios técnicos-científicos para o esforço contínuo da Comissão Internacional Baleeira (CIB) na avaliação do impacto da caça comercial e da recuperação das populações de baleias jubarte.

Os resultados da avaliação da estrutura populacional das baleias jubarte dos estoques reprodutivos Sul Americanos A e G (Capítulo 3) revelaram uma diferenciação significativa entre essas populações, em ambos marcadores moleculares (DNA mitocondrial e microssatélites), e especialmente através da análise bayesiana realizada pelo programa Structure, que identificou duas populações mesmo quando não foram informados os locais de amostragem. No entanto, os testes de *assignment* indicaram que um intercâmbio de indivíduos está ocorrendo entre essas populações, mas provavelmente com um fluxo gênico baixo o suficiente para permitir a independência demográfica desses dois estoques reprodutivos. Portanto, esses resultados indicam que as populações de baleias jubarte do Atlântico Sul Ocidental (A) e do Pacífico Sul Oriental (G) representam distintos estoques biológicos, baseados na separação genética (Donovan 1991), podendo assim ser considerados como distintas unidades de manejo, e suportando o reconhecimento desses dois estoques de jubartes do Hemisfério Sul definidos pela CIB.

Com relação aos dados separados por sexo, houve uma maior e significativa diferenciação entre as fêmeas do Brasil e da Colômbia, e uma maior segregação entre os machos do Brasil e da Península Antártica. Esses resultados sugerem que as fêmeas apresentam maior fidelidade às suas áreas de reprodução, e que os machos devem ser mais fiéis às áreas de alimentação, mas mais propensos a mudar seus destinos entre as áreas reprodutivas. No entanto, o intercâmbio entre fêmeas de diferentes áreas de reprodução também ocorre, como foi demonstrado por recentes estudos, os quais relataram fêmeas realizando longos movimentos entre áreas de reprodução em diferentes oceanos (Stevick et al. 2010, 2011). Isso indica uma flexibilidade comportamental para uma espécie usualmente filopátrica. Portanto, para uma melhor compreensão de como o intercâmbio está ocorrendo entre essas populações, e da variação do padrão migratório entre os sexos, é necessário um

esforço de amostragem principalmente na chegada e na saída das baleias jubarte migrando para essas áreas de reprodução.

Para uma melhor compreensão da história demográfica e com isso também do real impacto da caça comercial na população de baleias jubarte do estoque A (Capítulos 2 e 4), foram estimados seus tamanhos efetivos e de censo, contemporâneo e histórico, além das flutuações populacionais ao longo do tempo. Apesar da análise de computação Bayesiana aproximada (*Approximate Bayesian computacion* - ABC) ter suportado um cenário de população constante sobre os cenários de alteração do tamanho da população durante o período da caça comercial, as estimativas de tamanho (efetivo e de censo) atual da população em diferentes períodos de tempo demonstraram uma flutuação do tamanho da população durante esse período (~ 2 a 4 gerações atrás) (Capítulo 2). Esses resultados sugerem que o uso de diferentes métodos que estimam $N_e$ em diferentes períodos de tempo deve auxiliar na detecção de recentes mudanças demográficas das populações de baleias.

A estimativa de tamanho histórico da população indica que a população de baleias jubarte do Oceano Atlântico Sul Ociental (estoque A) foi muito maior (~ 6 vezes) do que aquele estimado pelos registros da caça (Capítulo 4), corroborando as recentes estimativas de grande tamanho histórico de população para as jubartes do Oceano Atlântico Norte, bem como para outras espécies de cetáceos (Alter *et al*. 2007, 2012, Alter & Palumbi 2009, Ruegg *et al*. 2013). Considerando que essa estimativa é a média ponderada do tamanho da população a longo prazo, ela tem sido influenciada pela dinâmica demográfica da população. Além disso, como os estoques reprodutivos do Oceano Atlântico Sul apresentam uma baixa divergência genética (Rosenbaum *et al*. 2009), essa estimativa deve provavelmente incluir a metapopulação de jubartes desse oceano.

Finalmente, a análise de *Extended Bayesian Skyline Plot* (EBSP) indicou pela primeira vez um declínio de população mais recente causado pela exploração antropogênica nos últimos 200 anos. Apesar de estudos anteriores para outras espécies de baleia não terem detectado um *bottleneck* mais recente usando *Bayesian Skyline Plot*, nosso resultado sugere que o maior número de locos (35 locos nucleares e região controle do DNAmt) usado neste estudo aumentou o poder estatístico para detecção de recentes declínios de população. O EBSP também apresentou um declínio da população iniciando a 30 mil anos atrás. Esse resultado sugere que a abundância de jubartes durante o final do Pleistoceno foi relacionada principalmente com as mudanças climáticas causadas pelos ciclos de glaciação/interglaciação, as quais geraram oscilações do nível do mar e flutuações das temperaturas dos oceanos, alterando assim a produtividade dos oceanos e consequentemente afetando as populações de

baleias. Além disso, esse resultado sugere que a população de baleias jubarte do Oceano Atlântico Sul Ocidental estava declinando antes do início da caça, o que deve explicar a discrepância encontrada entre as estimativas de tamanho de população, genéticas e baseadas nos dados da caça.

Os resultados obtidos nesta tese forneceram novas informações acerca da estrutura populacional e do intercâmbio entre as baleias jubarte da costa leste e oeste da América do Sul, respectivamente estoques reprodutivos A e G. Dado a significante diferenciação genética e o baixo nível de fluxo gênico entre essas populações, esses estoques reprodutivos (A e G) devem ser reconhecidos como distintas unidades de manejo pela CIB. Os resultados da tese também destacam a importância de levar em consideração a abundância histórica e os prováveis impactos ambientais que afetaram essa abundância para a elaboração dos planos de manejo para a conservação de uma espécie ameaçada e importante ecologicamente.

Para aprimorar as nossas estimativas genéticas do tamanho histórico da população de jubartes do Oceano Atlântico Sul Ocidental será necessário futuras investigações dos efeitos da estrutura populacional de jubartes no Oceano Atlântico Sul, além da redução de alguns fatores de incerteza, tais como as taxas de mutação, tempo de geração e relação entre tamanho de censo e efetivo para a espécie, os quais afetam a acuracidade dessas estimativas.

**Referências bibliográficas**

Acevedo-Gutiérrez A, Smultea MA. 1995. First records of humpback whales including calves at Golfo Dulce and Isla del Coco, Costa Rica, suggesting geographical overlap of northern and southern hemisphere populations. Mar. Mamm Sci. 11(4):554–60.

Acevedo J, Acuña P, Olavarría C, Aguayo-Lobo A, Pastene LA. 2004. Report of cetacean surveys in the Magellan Strait in the austral summer 2003/2004. Unpublished paper SC/56/O7 presented to the Scientific Committee of the International Whaling Commission.

Acevedo J, Rasmussen K, Félix F, Castro C, Llano M, Secchi ER, Saborio MT, Aguayo-Lobo A, Haase B, Scheidat M, Dalla-Rosa L, Olavarría C, Forestell P, Acuña P, Kaufman G, Pastene L. 2007. Migratory destination of humpback whales from the Magellan Strait feeding ground, Southeast Pacific. Mar Mamm Sci. 23:453-463.

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

Alter SE, Rynes E, Palumbi SR. 2007. DNA evidence for historic population size and past ecosystem impacts of gray whales. Proc Natl Acad Sci USA. 104:15162-15167.

Alter SE, Palumbi SR. 2009. Comparing evolutionary patterns and variability in the mitochondrial control region and cytochrome b in three species of baleen whales. J Mol Evol. 68:97-111.

Alter SE, Newsome SD, Palumbi SR. 2012. Prewhaling genetic diversity and population ecology in eastern pacific grey whales: insights from ancient DNA and stable isotopes. PLoS ONE. 7:e35039.

Amos B. 1996. Levels of genetic variability in cetacean populations have probably changed little as a result of human activities. Rep Int Whaling Comm. 46:657-658.

Andriolo A, Martins CCA, Engel MH, Pizzorno JL, Más-Rosa S, Freitas A, Morete ME, Kinas PG. 2006. The first aerial survey to estimate abundance of humpback whale (*Megaptera novaeangliae*) in the breeding ground off Brazil. J Cetacean Res Manage. 8:307-311.

Andriolo A, Kinas PG, Engel MH, Martins CCA. 2010. Humpback whale population estimates and distribution along the Brazilian breeding ground. Endang Spec Resear. 11:233-243.

Baker CS, Herman LM, Perry A, Lawton WS, Straley JM, Wolman AA, Kaufman GD, Winn HE, Hall JD, Reinke JM, Ostman J. 1986. Migratory movement and population structure of humpback whales (*Megaptera novaeangliae*) in the Central and Eastern North Pacific. Mar Ecol Prog Ser. 31:105-119.

Baker CS, Perry A, Bannister JL, Weinrich MT, Abernethy RB, Calambokidis J, Lien J, Lambertsen RH, Urban-Ramirez J, Vasquez O, Clapham PJ, Alling A, O´Brien SJ, Palumbi SR. 1993. Abundant mitochondrial DNA variation and worldwide population structure in humpback whales. Proc Natl Acad Sci USA. 90:8239-8243.

Baker CS, Slade RW, Bannister JL, Abernethy RB, Weinrich MT, Lien J, Urban-Ramirez J, Corkeron P, Calambokidis J, Vasquez O, Palumbi SR. 1994. Hierarchical structure of mitochondrial DNA gene flow among humpback whales, world-wide. Mol Ecol. 3:313-327.

Baker CS, Medrano-Gonzalez, Calambokidis J, Perry A, Pichler F, Rosenbaum H, Straley JM, Urban-Ramirez J, Yamaguchi M, von Ziegesar O. 1998. Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. Mol Ecol. 7:695-707.

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

Caballero S, Hamilton H, Jaramillo C, Capella J, Flórez-González L, Olavarría C, Rosenbaum HC, Guhl F, Baker CS. 2001. Genetic characterization of the Colombian pacific coast humpback whale population using RAPD and mitochondrial DNA sequences. Mem Quensl Mus. 47:459-464.

Calambokidis J, Steiger GH, Evenson JR, Flynn KR, Balcomb KC, Claridge DE, Bloedel P, Straley JM, Baker CS, von Ziegesar O, Dahlheim ME, Waite JM, Darling JD, Ellis G, Green GA. 1996. Interchange and isolation of humpback whales off California and other North Pacific feeding grounds. Mar Mamm Sci. 12:215-226.

Calambokidis J, Steiger GH, Rasmussen K, Urbán JR, Balcomb KC, Ladrón de Guevara P, Salinas MZ, Jacobsen JK, Baker CS, Herman LM, Cerchio S, Darling JD. 2000. Migratory destinations of humpback whales that feed off California, Oregon and Washington. Mar Ecol Prog Ser. 192:295-304.

Calambokidis J, Steiger GH, Straley JM, Herman LM, Cerchio S, Salden DR, Urbán J, Jacobson JK, von Ziegesar O, Balcomb KC, Gabrielle CM, Dahlheim ME, Uchida S, Ellis G, Miyamura Y, Ladrón de Guevara PP, Yamaguchi M, Sato F, Mizroch SA, Schlender L, Rasmussen K, Barlow J, Quinn TJ. 2001. Movements and population structure of humpback whales in the North Pacific. Mar Mamm Sci. 17:769-794.

Calambokidis J, Falcone EA, Quinn TJ, Burdin AM, Clapham PJ, Ford JKB, Gabriele CM, LeDuc R, Mattila D, RojasBracho L, Straley JM, Taylor BL, Urbán J, Weller D, Witteveen BH, Yamaguchi M, Bendlin A, Camacho D, Flynn K, Havron A, Huggins J,

Malone N, Barlow J, Wade PR. 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).

Capella J, Gibbons J, Flórez-González I, Llano M, Valladares C, Sabaj V, Vilina Y. 2008. Migratory round-trip of individually identified humpback whales at the Strait of Magellan: clues on transit times and phylopatry to destinations. Revista Chilena de Historia Natural. 81:547-60.

Chittleborough RG. 1965. Dynamics of two populations of the humpback whale, Megaptera novaeangllae (Borowski). Aust J Mar Freshw Res. 16:33-128.

Clapham PJ, Mead JG. 1999. *Megaptera novaeangliae*. Mammalian Species. 604:1-9.

Clapham PJ, Baker CS. 2002. Modern whaling. In: Perrin WF, Würsig B, Thewissen JGM, editors. Encyclopedia of Marine Mammals, New York: Academic Press; p. 1328-1332.

Constantine R, Russell K, Gibbs N, Childerhouse S, Baker CS. 2007. Photo-identification of humpback whales (*Megaptera novaeangliae*) in New Zealand waters and their migratory connections to breeding grounds of Oceania. Mar Mamm Sci. 23(3):715-720.

Cypriano-Souza AL, Fernández GP, Lima-Rosa CAV, Engel MH, Bonatto SL. 2010. Microsatellite genetic characterization of the humpback whale (*Megaptera novaeangliae*) Breeding Ground off Brazil (Breeding Stock A). J Hered. 101(2):189-200.

Darling JD, Cerchio S. 1993. Movement of a humpback whale, *Megaptera novaeangliae*, between Japan and Hawaii. Mar Mamm Sci. 9:84-89.

Dawbin, W. 1966. The seasonal migratory cycle of the humpback whale. In: *Whales, Dolphins and Porpoises.* (ed. Norris KS), pp 145-170. University of California Press, Berkeley.

Donovan GP. 1991. A review of IWC stock boundaries. In: Hoelzel AR, editor. Genetic ecology of whales and dolphins. Rep Int Whaling Comm Spec Iss. 13:39-68.

Engel MH, Fagundes NJR, Rosenbaum HC, Leslie MS, Ott PH, Schmitt R, Secchi E, Dalla Rosa L, Bonatto SL. 2008. Mitochondrial DNA diversity of the Southwestern Atlantic humpback whale (*Megaptera novaeangliae*) breeding area off Brazil, and the potential connections to Antartic feeding areas. Conserv Genet. 9:1253-1262.

Engel MH, Martin AR. 2009. Feeding grounds of the western South Atlantic humpback whale population. Mar. Mamm. Sci. 25:964–969.

Félix F, Haase B. 2005. Distribution of humpback whales along the coast of Ecuador and management implications. J Cetacean Res Manage. 7(1):21-31.

Félix F, Palacios DM, Caballero S, Haase B, Falconí J. 2006. The 2005 Galápagos Humpback Whale Expedition: a first attempt to assess and characterize the population in the Archipelago. Unpublished paper SC/A06/HW15 presented to the Scientific Committee of the International Whaling Commission.

Félix F, Rasmussen MH, Garita F, Haase B, Simonis A. 2009. Movements of humpback whales between Ecuador and central America, wintering area of the breeding stock G. Unpublished paper SC/61/SH18 presented to the Scientific Committee of the International Whaling Commission.

Félix F, Caballero S, Olavarría C. 2012. Genetic diversity and population structure of humpback whales (*Megaptera novaeanglieae*) from Ecuador based on mitochondrial DNA analyses. J Cetacean Res Manage. 12(1):71-77.

Findlay KP. 2001. A review of humpback whale catches by modern whaling operations in the Southern Hemisphere. Mem Quensl Mus. 47(2):411-20.

Flórez-González L, Capella J, Haase B, Bravo GA, Félix F, Gerrodette T. 1998. Changes in winter destinations and the northernmost record of southeastern Pacific humpback whales. Mar. Mamm Sci.14(1):189-96.

Freitas AC, Kinas PG, Martins CCA , Engel MH. 2004. Abundance of humpback whales on the Abrolhos Bank wintering ground, Brazil. J Cetacean Res Manage. 3:225-230.

Furtado-Neto MA, Monteiro-Neto C, Campos A, Lien J, Carr S. 1998. Are Northern-Hemisphere humpback whales stranding in South Atlantic beaches? Answers from mitochondrial DNA sequences. Abstracts of the 8º Reunião de Trabalhos de Especialistas em Mamíferos Aquáticos da América do Sul and 2º Congresso da Sociedade Latino Americana de Mamíferos Aquáticos, SOLAMAC.

Garrigue C, Aguayo-Lobo A, Amante-Helwing VLU. 2002. Movements of humpback whales in Oceania, South Pacific. J Cetacean Res Manage. 4:255-260.

Gibbons J, Capella JJ, Valladares C. 2003. Rediscovery of a humpback whale (*Megaptera novaeangliae*) feeding ground in the Straits of Magellan, Chile. J Cetacean Res Manage. 5:203-208.

Hoelzel AR. 1991. Genetic ecology of whales and dolphins. Report of the International Whaling Commission (Special Issue 13), Vol. International Whaling Commission, Cambridge.

Hoelzel AR. 1998. Genetic structure of cetacean populations in sympatry, parapatry, and mixed assemblages: Implications for conservation policy. J Hered. 89:451-458.

IBAMA, 2001. *Mamíferos Aquáticos do Brasil: Plano de Ação*, versão 2, 2ª ed., Brasília.

International Whaling Commission, 2005. Report of the Scientific Committee. Annex H. Report of the sub-committee on other Southern Hemisphere whale stocks. J Cetacean Res Manage. 7:235-46.

IUCN2008. 2008 IUCN red list of threatened species. [Internet]. [cited 4 May 2009]. Available from: http://www.iucnredlist.org/.

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.

Jann B, Allen J, Carrillo M, Hanquet S, Katona SK, Martin AR, Reeves RR, Seton R, Stevick PT, Wenzel FW. 2003. Migration of a humpback whale (*Megaptera novaeangliae*) between the Cape Verde Islands and Iceland. J Cetacean Res Manage. 5:125-129.

Lodi L. 1992. Uma história da caça à baleia. Ciência Hoje. 14(81):78-83.

Katona SK, Whitehead H. 1981. Identifying humpback whales using their natural marks. Polar Record. 20:439-444.

Kellogg R. 1929. What is known of the migration of some of the whalebone whales. Smithsonian Institution Annual Report. 1928:467-494.

Lodi L. 1994. Ocorrências de baleias-jubarte, *Megaptera novaeangliae*, no Arquipélago de Fernando de Noronha, incluindo um resumo de registros de capturas no Nordeste do Brasil. Biotemas. 7:116-123.

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

Mackintosh NA. 1946. The natural history of whalebone whales. Biological Reviews. 21(2):60-74.

Mackintosh NA. 1965. The stocks of whales. Vol. Fishing News (Books) Ltd., London.

Magalhães FA, Tosi CH, Garri RG, Chellappa S, Silva FL. 2008. Cetacean diversity on the Parnaiba Delta, Maranhão state, northeastern Brazil. Braz J Biol. 68(3):545-551.

Martins CCA, Morete ME, Engel MH, Freitas A, Secchi ER, Kinas PG. 2001. Aspects of habitat use patterns of humpback whales in the Abrolhos bank, Brazil, breeding ground. Mem Queensl Mus. 47:563-570.

Mattila DK, Clapham PJ, Vasquez O, Bowman RS. 1994. Occurrence, population composition, and habitat use of humpback whales in Samana Bay, Dominican Republic. Canadian J Zool. 72:1898-1907.

Medrano-González L, Baker CS, Robles-Saavedra MR, Murrell J, Vazquez-Cuevas MJ, Congdon BC, Stranley JM, Calambokidis J, Urban-Ramirez J, Flórez-González L, Olavarría C, Aguayo-Lobo A, Nolasco-Soto J, Juarez-Salas RA, Villavicencio-Llamosas

K. 2001. Trans-oceanic population genetics structure of humpback whales in the North and South Pacific. Mem Quensl Mus. 47:465-479.

Meirelles ANO, Monteiro-Neto C, Martins AMA, Costa AF, Barros HMDR, Alves MDO. 2009. Cetacean strandings on the coast of Ceará, northeastern Brazil (1992-2005). J Mar Biol Assoc United Kingdom. 89(5):1083-1090.

Mikhalev YA. 1997. Humpback whales *Megaptera novaeangliae* in the Arabian Sea. Mar Ecol Prog Ser. 149:13-21.

Olavarría C, Baker CS, Medrano L, Aguayo A, Caballero S, Flórez-González L, Capella J, Rosenbaum HC, Garrigue C, Greaves J, Jenner M, Jenner C, Bannister JL. 2000. Stock identity of Antarctic Peninsula Humpback whales inferred from mtDNA variation. Unpublished paper SC/52/IA15 presented to the Scientific Committee of the International Whaling Commission.

Olavarría, C., Aguayo, A., Acevedo, J., Medrano, L., Thiele, D. and Baker, C.S. 2006. Genetic differentiation between two feeding areas of the Eastern South Pacific humpback whale population: update on SC/57/SH3. Unpublished paper SC/A06/HW29 presented to the Scientific Committee of the International Whaling Commission.

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 humpback whales throughout the South Pacific and the origin of the eastern Polynesian breeding grounds. Mar Ecol Prog Ser. 330:257-268.

Pacheco AS, Silva S, Alcorta B. 2009. Winter distribution and group composition of humpback whales (*Megaptera novaeangliae*) off Northern Peru. Lat Am J Aquat Mamm. 7(1-2):33-38.

Paiva MP, Grangeiro BF. 1965. Biological investigations on the whaling seasons 1960-1963, off the northeastern coast of Brazil. Arq Estac Biol Mar Univ Ceara. 5:29-64.

Paiva MP, Grangeiro BF. 1970. Investigations on the whaling seasons 1964-1967, off the northeastern coast of Brazil. Arquivos de Ciências do Mar. 10:111-126.

Palsbøll PJ, Clapham PJ, Mattila DK, Larsen F, Sears R, Siegismund HR, Sigurjónsson J, Vásquez O, Artander P. 1995. Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. Mar Ecol Prog Ser. 116:1-10.

Palsbøll PJ, Allen J, Berube M, Clapham PJ, Feddersen TP, Hammond PS, Hudson RR, Jørgensen H, Katona S, Larsen AH, Larsen F, Lien J, Mattila DK, Sigurjonsson J, Sears

R, Smith T, Sponer R, Stevick P, Øien N. 1997. Genetic Tagging of Humpback Whales. Nature. 388:767-769.

Pinedo MC. 1985. A note on a stranding of the humpback whale on the southern coast of Brazil. Sci. Rep. Whales Res. Inst. 36:165-8.

Pretto DJ, Andrade MCM, Oliveira JM, Oliveira MGA. 2009. First record of a humpback whale, *Megaptera novaeangliae* (Borowski, 1781), stranding in Pará State, Northern coast of Brazil. Braz J Biol. 69(4):1207-1208.

Rasmussen K, Palacios D, Calambokidis J, Saborio MT, Dalla Rosa L, Secchi ER, Steiger GH, Allen JM, Stone G. 2007. Southern Hemisphere humpback whales wintering off Central America: insights from water temperature into the longest mammalian migration. Biology Letters. 3(3):302-05.

Rayner GW. 1940. Whale marking. Progress and results to December 1939. Discovery Reports. 19:245-284.

Reeves RR, Clapham PJ, Wetmore SE. 2002. Humpback whale (*Megaptera novaeangliae*) occurrence near the Cape Verde Islands, based on American 19th century whaling records. J Cetacean Res Manage. 4:235-253.

Roman J, Palumbi S. 2003. Whales before whaling in North Atlantic. Science. 301:508-510.

Rooney AP, Honeycutt RL, Derr JN. 2001. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. Evolution. 55:1678-1685.

Rosenbaum HC, Pomilla C, Mendez M, Leslie MS, Best PB, Findlay KP, Minton G, Ersts PJ, Collins T, Engel MH, Bonatto SL, Kotze DPGH, Meÿer M, Barendse J, Thornton M, Razafindrakoto Y, Ngouessono S, Vely M, Kiszka J. 2009. Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. PLoS ONE 4:e7318.

Rossi-Santos MR, Neto ES, Baracho CG, Cipolotti SR, Marcovaldi E, Engel MH. 2008. Occurrence and distribution of humpback whales (Megaptera novaeangliae) on the North coast of the State of Bahia, Brazil, 2000-2006. ICES J Mar Sci. 65:1-7.

Ruegg KC, Anderson EC, Baker CS, Vant M, Jackson JA, Palumbi SR. 2010. Are Antarctic minke whales unusually abundant because of 20th century whaling? Mol Ecol. 19:281-291.

Ruegg KC, Rosenbaum HC, Anderson EC, Engel MH, Rothchild A, Baker CS, Palumbi SR. 2013. Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure. Conserv Genet. 14:103-114.

Siciliano S. 1997. Características da população de baleias-jubarte (*Megaptera novaeangliae*) da costa brasileira, com especial referência aos Bancos de Abrolhos. MS Thesis, Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro.

Siciliano S, Pizzorno JLA, Barata PCR. 1999. Distribution and possible migratory routes of humpback whales *Megaptera novaeangliae* in the western South Atlantic. Unpublished paper SC/51/CAWS4 to the Scientific Committee of the International Whaling Commission.

Siciliano S, Emin-Lima NR, Costa AF, Rodrigues ALF, Magalhães FA, Tosi CH, Garri RG, da Silva CR, Silva Jr JS. 2008. Revisão do conhecimento sobre os mamíferos aquáticos da costa norte do Brasil. Arquivos do Museu Nacional, Rio de Janeiro. 66(2):381-401.

Siciliano S, Moura JF, Filgueiras HR, Rodrigues PP, Leite Jr. NO. 2012. Sightings of humpback whales on the Vitória-Trindade Chain and around Trindade Island, Brazil. Braz J Oceano. 60(3): 455-459.

Slijper EJ. 1962. Whales. 1st English Edn. Hutchinson & Co., London. 475pp.

Slijper EJ. 1965. A hundred years of modern whaling. Netherlands Commmission for International Nature Protection, Amsterdam. 43pp.

Stevick PT, Aguayo A, Allen J, Avila IC, Capella J, Castro C, Chater K, Engel M, Félix F, Flórez-González L, Freitas A, Hasse B, Llano M, Lodi L, Muñoz E, Olavarría C, Secchi E, Scheidat M, Siciliano S. 2004. A note on the migrations of individually identified humpback whales between the Antarctic Peninsula and South America. J Cetacean Res Manag. 6:109–113.

Stevick PT, Allen J, Clapham PJ, Katona SK, Larsen F, Lien J, Mattila DK, Sears PJPR, Sigurjónsson J, Smith TD, Vikingsson G, Øien N, Hammond PS. 2006a. Population spatial structuring on the feeding grounds in North Atlantic humpback whales (*Megaptera novaeangliae*). J Zool. 270:244-255.

Stevick PT, Pacheco de Godoy L, McOsker M, Engel MH, Allen A. 2006b. A note on the movement of a humpback whale from Abrolhos Bank, Brazil to South Georgia. J Cetacean Res Manag. 8:297–300.

Stevick PT, Neves MC, Johansen F, Engel MH, Allen J, Marcondes MCC, Carlson C. 2010. A quarter of a world away: female humpback whale moves 10,000 km between breeding areas. Biology Letters. 7:299-302.

Stevick PT, Allen JM, Engel MH, Félix F, Haase B, Neves M. 2011. First record of inter-oceanic movement of a humpback whale between Atlantic and Pacific breeding grounds

off South America. Report SC/63/SH4 to the Scientific Committee of the International Whaling Commission.

Stone GS, Flórez-González L, Katona S. 1990. Whale migration record. Nature. 346: 705-706.

Tønnessen JN, Johnsen AO. 1982. The history of modern whaling. Berkeley (CA): University of California Press.

Urbán J, Jaramillo A, Aguayo A, Ladrón de Guevara P, Salinas M, Alvarez C, Medrano L, Jacobsen JK, Balcomb KC, Claridge DE, Calambokidis J, Steiger GH, Straley JM, Ziegesar Ov, Waite JM, Mizroch S, Dahlheim ME, Darling JD, Baker CS. 2000. Migratory destinations of humpback whales wintering in the Mexican Pacific. J Cet Res Manage. 2:101-110.

Valsecchi E, Corkeron PJ, Galli P, Sherwin W, Bertorelle G. 2010. Genetic evidence for sex-specific migratory behaviour in western South Pacific humpback whales. Mar Ecol Prog Ser. 398:275-86.

Waldick RC, Kraus S, Brown M, White BN. 2002. Evaluating the effects of historic bottleneck events: an assessment of microsatellite variability in the endagered, North Atlantic right whale. Mol Ecol. 11:2241-2249.

Ward E, Zerbini, AN, Kinas PG, Engel MH, Andriolo A. 2011. Estimates of population growth rates of humpback whales (*Megaptera novaeangliae*) in the wintering grounds off the coast of Brazil (Breeding Stock A). J Cetacean Res Manage. 3:145-149.

Wedekin LL, Engel MH, Azevedo A, Kinas PG, Andriolo A, Simões-Lopes PC. 2010. Density and abundance of the humpback whale in the Brazilian breeding ground (stock A): aerial survey, 2008. Report SC/62/SH28 presented to the Scientific Committee of the International Whaling Commission.

Wedekin LL, Rossi-Santos M, Baracho C, Cypriano-Souza AL, Simões-Lopes PC. in press. Cetacean records along a coastal-offshore gradient in the Vitória-Trindade Chain, western South Atlantic Ocean. Braz J Biol.

Williamson GR. 1975. Minke whales off Brazil. Scientific Reports of the Whales Research Institute. 27:37-59.

Yablokov AV, Zemsky VA, Mikhalev YA, Tornosov VV, Berzin AA. 1998. Data on Soviet whaling in the Antarctic in 1947-1972 (population aspects). Russ J Ecol. 29:38-42.

Zerbini AN, Andriolo A, da Rocha JM, Simões-Lopes PC, Siciliano S, Pizzorno JL, Waite JM, DeMaster DP, VanBlaricon GR. 2004. Winter distribution and abundance of

humpback whales (*Megaptera novaeangliae*) off Northeastern Brazil. J. Cetacean Res. Manage, 6, 101-107.

Zerbini AN, Andriolo A, Heide-Jorgensen MP, Pizzorno JL, Maia YG, VanBlaricom GR, DeMaster DP, Simões-Lopes PC, Moreira S, Bethlem C. 2006. Satellite-monitored movements of humpback whales *Megaptera novaeangliae* in the Southwest Atlantic Ocean. Mar Ecol Prog Ser. 313:295-304.

Zerbini AN, Andriolo A, Danilewicz D, Heide-Jorgensen MP, Gales N, Clapham PJ. 2011. An update on research on migratory routes and feeding destinations of Southwest Atlantic humpback whales. Unpublished paper SC/63/SH23 presented to the Scientific Committee of the International Whaling Commission.