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

Uma abordagem genômica no estudo da história demográfica da população de baleias jubarte (*Megaptera novaeangliae*) do Atlântico sul ocidental

Júlia Dworakowski Emerim

DISSERTAÇÃO DE MESTRADO

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
Av. Ipiranga 6681 - Caixa Postal 1429
Fone: (051) 320-3500
CEP 90619-900 Porto Alegre - RS
Brasil

2016

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

Uma abordagem genômica no estudo da história demográfica da população de baleias jubarte (*Megaptera novaeangliae*) do Atlântico sul ocidental

Júlia Dworakowski Emerim
Orientador: Prof. Dr. Sandro Luis Bonatto

DISSERTAÇÃO DE MESTRADO
PORTO ALEGRE – RS – BRASIL
2016

SUMÁRIO

AGRADECIMENTOS

Inicio meus agradecimentos dizendo que esta foi uma trajetória importante pra mim e que muitas pessoas fizeram parte dela e também foram importantes. Espero conseguir prestar meus agradecimentos a todos os envolvidos da forma mais justa possível.

Primeiramente gostaria de agradecer ao meu orientador Sandro Bonatto, pela oportunidade e confiança. Agradeço toda a sua orientação, todas as discussões, opiniões, sugestões, críticas (positivas e negativas) que contribuíram para o desenvolvimento desde projeto e para meu crescimento e aprendizado.

Não posso esquecer de mencionar a CAPES, que proporcionou uma bolsa de estudo para a realização desse.

A toda a minha família, em especial meus pais, que me deram apoio e aguentaram minhas crises, dias de mau-humor, dias eufóricos, ansiedade e nervosismo durante esse período. Eles que apesar de tudo, e apesar de nada, estavam ao meu lado.

Gostaria de agradecer também a Fernanda Pedone, nossa laboratorista, pelas horas de envolvimento com o projeto, discussões de protocolo e pesquisa para conseguirmos superar as dificuldades que surgiram durante o percurso. Da mesma maneira à Cristine Trinca, que cedeu algumas de suas horas para concretizarmos as últimas etapas do protocolo de bancada. Assim como as colegas Giovanna Giudicelli e Maikel Kortmann.

A colega Ana Lúcia Cypriano Souza pelo auxílio com a seleção de amostras.

Ao colega de laboratório e amigo Manuel, que com muita paciência me ajudou com muitas de minhas dúvidas e principalmente questões de bioinformática. Mas que também participou de muitos momentos bons, junto com a Letícia e também colegas e amigos Fernando, Fabrício, Isadora, Maísa, e Tiago que me ouviram e aguentaram meus dramas. Agradeço a eles o companheirismo nos bons e maus momentos, nos cursos, nas aulas e trabalhos, nos almoços e as idas ao RU, essa última nem tanto.

Ao Gustavo Arias, que no último mês cedeu parte de seu tempo para que algumas das análises pudessem ser finalizadas.

À minha prima Olenka que cedeu uns minutinhos para me ajudar com a revisão ortográfica.

À Fernanda Trindade, Sara Santos, Henrique Figueiró, Maíra Rodrigues, agradeço pelas conversas, pelo apoio, troca de ideias e por se manterem disponíveis ao que pudessem ajudar e pela eventual ajuda é claro.

Agradeço aos meus amigos por todo apoio e incentivo, em especial a Tábata e Rogério, pela parceria, saídas e desabafos; e tantos outros que participaram mesmo que indiretamente nessa trajetória.

Não posso me esquecer de mencionar meus amigos e colegas circenses (Martina, Guilherme, Isabela, Consuelo, Paola 'Lola', Carol e Mariana 'Mêmis') que me proporcionaram ótimos momentos de reflexão, descontração, foco, criação, trabalho e diversão.

Resumo

Atualmente são reconhecidos sete estoques reprodutivos (A-G) de baleias jubarte (*Megaptera novaeangliae*) no hemisfério sul. O estoque reprodutivo 'A' estende-se ao longo costa brasileira (entre 5° e 23° S) sendo o Banco de Abrolhos-BA a principal área de reprodução do oceano atlântico sul para a espécie. Durante o período de caça baleeira comercial (início do século XX) sugere-se que esse estoque tenha sofrido um grande impacto populacional chegando a aproximadamente 2% do seu tamanho original. Pesquisas recentes, utilizando diferentes técnicas para estudo da história demográfica, têm apresentado controvérsias acerca dos valores de abundância recente e histórico para essa população. Para o delineamento de estratégias de conservação é de grande importância o conhecimento a diversidade genética e as possíveis flutuações populacionais ao longo do tempo e em determinados períodos. No presente estudo fez-se o sequenciamento de 25 amostras de DNA extraídas de tecido de indivíduos da população Brasileira de baleias jubarte, através da construção de bibliotecas de ddRADseq e sequenciamento de nova geração. Os dados evidenciam a ausência de estruturação populacional na população. A análise com 5145 locos com o migrate-n estimou a diversidade genética desta população em 0.00237 por sítio, que indica um tamanho efetivo de aproximadamente 40.000 e tamanho censitário de ~140.000. A abordagem de ABC (*Approximate Bayesian Computation*), usando ~mil locos nucleares, aplicada para testar diferentes cenários demográficos relacionados ao período de caça apontou para um cenário de população recente (<10 gerações) constante (contra cenários com alterações populacionais neste período) corroborando estudos prévios utilizando microssatélites e poucos loci nucleares. A estimativa para o Ne no cenário constante foi semelhante ao obtido com o migrate-n. Por fim, o skyline plot obtido com o migrate-n sugere um aumento do tamanho efetivo de mais de uma ordem de magnitude se estendendo por centenas de milhares de geração, o que sugere que esta estimativa pode na verdade refletir o tamanho populacional de uma da metapopulação abrangendo todo o hemisfério sul ou mesmo toda a espécie, hipótese que precisa ser testada com a adição de outras populações e cenários demográficos apropriados.

Abstract

A genomic approach to the study of the demographic history of the Southwestern Atlantic humpback whale (Megaptera novaeangliae) population

Currently there are seven recognized reproductive stocks (A-G) of humpback whales (*Megaptera novaeangliae*) in the southern hemisphere. The breeding stock 'A' spreads along the Brazilian coast (between 5 ° and 23 ° S) and the Abrolhos Bank- BA is the main breeding area of the Southwest Atlantic Ocean. During the commercial whaling period (early 20th century) the breeding stock 'A' had reached nearly of 2% of its historical size. Recent researches, using different techniques to study the demographic history, have found different values for recent and historical abundance for this population. The knowledge about population size at specific times and its dynamics during time is very important to draw conservations strategies. Here we sequenced a ddRADseq library of 25 DNA samples extracted from tissues of individuals from the Brazilian humpback whales population. The data suggests absence of population structure in the population. The analysis with 5145 locus with migrate-n estimated the genetic diversity of the population as 0.00237 in per site, indicating an effective size of approximately 40,000 and a census size ~140,000. The ABC approach (Approximate Bayesian Computation), used to test different demographic scenarios related to the commercial whaling period, supported a constant population scenario (<10 generations) (against scenarios with population changes in this period) corroborating previous studies using microsatellites and a few nuclear loci. The Ne estimated in the constant scenario was similar to that obtained with the migrate-n method. Finally, the skyline plot obtained with the migrate-n suggests an increase in the effective size of more than an order of magnitude extending for hundreds of thousands of generation in the past. This very long time frame suggests that this estimate may actually reflects the population size of a metapopulation covering the entire Southern Hemisphere or even entire species, a hypothesis that needs to be tested with the addition of other populations and appropriate demographic scenarios.

.

Apresentação

A presente dissertação foi desenvolvida como parte dos requisitos para obtenção do título de Mestre pelo Programa de Pós-Graduação em Zoologia, da Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul.

Nesta pretendeu-se inferir a história demográfica da população Brasileira de baleias jubarte (*Megaptera novaeangliae* Borowski, 1871) baseado em analises genômicas de 25 indivíduos do Banco de Abrolhos, BA. A baleia jubarte, espécie cosmopolita (Dawbin 1966), pertence à família Balaenopteridae (Johnson and Wolman 1984). Essas têm hábito migratório, percorrem extensas distancias desde as regiões polares e sub-polares até o Equador (Dawbin 1966). No hemisfério sul as áreas de alimentação podem ser divididas em seis (I - VI), e as áreas de reprodução, também chamadas de estoques, em sete (A-G) (IWC – *International Whaling Commission*, 2005). A população em estudo configura o estoque reprodutivo A, e alimenta-se na porção leste da área II perto das ilhas Geórgias do Sul e Sanduiche do Sul (Engel et al. 2008, Engel and Martin 2009, Stevick et al. 2006, Zerbini et al. 2006a,).

O começo do século XX foi marco do início da caça baleeira comercial, logo, da ampliação das atividades das estações baleeiras (Clapham and Baker 2002, Findlay 2001), e considera-se motivo da diminuição da população mundial de baleias jubarte. Em 1982 a Comissão Internacional Baleeira (CIB) previu uma pausa da caça comercial de baleias em todos os estoques a partir das temporadas de 1985/1986, ainda em vigor visto que a população mundial de jubartes em 1966 já apresentava uma redução para menos de 10% da original (Tonnessen and Johnsen 1982). Estima-se, por registro de capturas, que o tamanho populacional da espécie na costa ocidental da América do Sul antes da caça comercial baleeira era de aproximadamente 24.700 indivíduos (Zerbini et al. 2006b). Em contraste, estudos mais recentes indicam valores de abundância histórica pré-caça de 4 a 6 vezes maior, esses utilizando estimativas SMM (*stepwise mutation model*) (Cypriano-Souza et al. 2014) e multi-locus (Cypriano-Souza 2013).

A baleia jubarte é uma espécie de grande importância para a conservação da biodiversidade, no entanto ainda existe muito desconhecimento e controvérsias científicas acerca de parâmetros fundamentais para o desenho de estratégias de conservação, tais como o tamanho populacional nos diferentes estoques reprodutivos antes do início da caça. O desconhecimento é ainda maior sobre a história evolutiva, incluindo a dinâmica demográfica histórica.

O genoma *M. novaeangliae* apresenta alta diversidade em diversos *loci* nucleares (Cypriano-Souza et al. 2014), logo contém, muita informação genética que pode ser utilizada para a reconstrução demográfica da espécie (Jackson et al. 2008), incluindo eventos bem antigos.

Para a descoberta de variabilidade molecular em populações naturais têm sido desenvolvidas várias estratégias de NGS, podendo-se citar as "*reduced representation libraries*" (RRLs) como uma das principais, a qual vem sendo utilizada com eficiência em espécies com genomas desconhecidos (ex. Garvin et al. 2010, Maughan et al. 2009). Nesse contexto utilizou-se uma técnica conhecida como *Double Digest Restriction Associated DNA* (ddRAD) *Sequencing*, seguindo o protocolo descrito por DaCosta and Sorenson (2014), com a qual conseguiu-se obter um grande volume de marcadores.

Os dados apresentados foram gerados a partir de amostras de tecido coletadas no ano de 2012 da população brasileira de baleias jubarte, fornecidas pelo Instituto Baleia Jubarte. As bibliotecas genômicas de ddRADseq foram construídas no Laboratório de Biologia Genômica e Molecular (Faculdade de Biociências da PUCRS) bem como as análises subsequentes, após sequenciamento na plataforma Illumina.

A dissertação será apresentada no formato de artigo científico a ser submetido à revista Journal of Heredity, respeitando as normas de submissão da mesma, disponíveis em: http://jhered.oxfordjournals.org/.

A genomic approach to the study of the demographic history of the Southwestern Atlantic humpback whale (*Megaptera novaeangliae*) population

Júlia D. Emerim[1], Manuel A.R. Escalona[1], Gustavo A.S. Arias[1], Marcia H. Engel[2], Sandro L. Bonatto[1]

[1]Faculdade de Biociências, PUCRS, Av. Ipiranga 6681, prédio 12C. Porto Alegre, RS 90619-900, 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, E-mail: slbonatto@pucrs.br

Abstract

Currently there are seven recognized reproductive stocks (A-G) of humpback whales (*Megaptera novaeangliae*) in the southern hemisphere. The breeding stock 'A' spreads along the Brazilian coast (between 5 ° and 23 ° S) and the Abrolhos Bank- BA is the main breeding area of the Southwest Atlantic Ocean. During the commercial whaling period (early 20th century) the breeding stock 'A' had reached nearly of 2% of its historical size. Recent researches, using different techniques to study the demographic history, have found different values for recent and historical abundance for this population. The knowledge about population size at specific times and its dynamics during time is very important to draw conservations strategies. Here we sequenced a ddRADseq library of 25 DNA samples extracted from tissues of individuals from the Brazilian humpback whales population. The data suggests absence of population structure in the population. The analysis with 5145 locus with migrate-n estimated the genetic diversity of the population as 0.00237 in per site, indicating an effective size of approximately 40,000 and a census size ~140,000. The ABC approach (Approximate Bayesian Computation), used to test different demographic scenarios related to the commercial whaling period, supported a constant population scenario (<10 generations) (against scenarios with population changes in this period) corroborating previous studies using microsatellites and a few nuclear loci. The Ne estimated in the constant scenario was similar to that obtained with the migrate-n method. Finally, the skyline plot obtained with the migrate-n suggests an increase in the effective size of more than an order of magnitude extending for hundreds of thousands of generation in the past. This very long time frame suggests that this estimate may actually reflects the population size of a metapopulation covering the entire Southern Hemisphere or even entire species, a hypothesis that needs to be tested with the addition of other populations and appropriate demographic scenarios.

Keywords: BSA, commercial whaling, demography, NGS, nuclear loci

Introduction

Humpback whales (*Megaptera novaeangliae Borowski*, 1781) have a seasonal migratory habit, travelling long distances from tropical waters to polar regions (Dawbin 1966). In winter they migrate to warm waters, in low latitudes, to conserve energy, calving and feed (Dawbin 1996, IUCN), the lower abundance of predators could explain why they migrate to these areas too (Corkeron and Connor 1999, Payne 1995). In summer, they migrate to higher latitudes, to their feeding areas, where there is abundance of krill (main component of their diet) (Brodie 1975). In the Southern Hemisphere, there are seven (termed A-G) breeding areas, also called stocks, and six (I-VI) feeding areas (International Whaling Commission -IWC). The breeding stock A (BSA), the main ground in western South Atlantic for breeding and calving for the species is situated specially around the Abrolhos Bank, Brazil (16°40'- 19°30' S and 37°25'- 39°45' W) (Engel 1996). The Brazilian humpback whale population has their feeding area in the eastern part of the area II, near South Georgia and South Sandwich (see Engel et al. 2008), that was also one of the main area of industrial modern whaling in Southern Hemisphere (Tonnessen and Johnsen 1982).

Commercial whaling initiated around the early 20th century and the expansion of the activities of the whaling stations resulted in the killing of thousands of whales. Over six decades about 2,000,000 large whales were killed throughout the Southern Hemisphere in general and around Antarctica feeding areas about 200,000 humpback whales were killed (Clapham and Baker 2002, Findlay 2001). A facilitating factor for the modern whaling methods is the coastal habit of this species, making it more easy to catch (e.g. Best 1994, Chittleborough 1965, Gambell 1973, Tonnessen and Johnsen 1982, Williamson 1975). Around the decade of 1950, the BSA had reached as low as 2% of its historical size, about 500 individuals (Zerbini et al. 2006). The IWC suspended the commercial whaling since 1985/1986 seasons in all stocks, considering the overexploitation of these populations, and in Brazil, the whaling is prohibited ever since 1987.

After the commercial whaling moratorium, the humpback whale population has increased steadily in the majority of the breeding areas, so that in 2008 the IUCN (International Union for the Conservation of Nature and Natural Resources) reclassified the species, updating its status to "least concern". According to Ward et al. (2011) the Brazilian humpback whale population has shown a growth of 7.4% per year between 1995 and 1998. The most recent studies of contemporary abundance of this population based on aerial surveys, estimated a number of individuals ranging between 6,404 (Andriolo et al. 2010) and 10,160 (Julião 2013). Estimates of the historical size of this population before whaling show a large discrepancy depending on the method used. Estimates using catch records suggested the population size was around 24,700 individuals (Zerbini et al. 2006). However, population size estimates based on genetic diversity reported for some baleen whale species (e.g. Alter et al. 2007, 2012, Roman and Palumbi 2003, Ruegg et al. 2010, 2013) are much higher than abundance estimates from catch records (Punt et al. 2006, Wade and Perryman 2002). Bad estimates from both kinds of methods does not seems to explain these consistent differences, as both approaches have highly improved their methods and the different persists. On the other hand, as previous genetic data provided long-term mean estimates (over hundreds or thousands of generations) rather than the estimate for the population right before the beginning of the whaling (Palsbøll et al. 2013), it is possible that the population sizes of whales just before whaling have been lower than their long-term sizes (Alter et al. 2007, 2012, Ruegg et al. 2013). Therefore, it would be necessary to better estimate the population size fluctuation over time, in special the recent effective size, if different from the long-term size. For these it would be important to increase significantly the number of genetic markers and use methods that could estimate the dynamics of the size of a population. Here we used the ddRAD sequencing approach (DaCosta and Sorenson 2014) to obtain sequence information of thousands of loci for the Southwestern Atlantic humpback whale population aiming to better estimate the genetic diversity of the population over the whole genome as well as to investigate its effective population size history. As for our knowledge, this is the first study of this kind in a true whale species.

Material and Methods

Sampling and DNA extraction

Tissue samples were collected using biopsy dart procedure (Lambertsen 1987) by a research team of the Humpback Whale Institute in the year 2012 at BSA breeding area in Brazil. DNA was extracted from these samples using DNeasy Blood and Tissue Kit (QIAGEN) and conserved in alcohol 70% at -20ºC. We selected 30 samples which had previously presented good genotyping results and are not closely related (Cypriano-Souza et al. 2010). DNA Quality was checked in 1% agarose gel and quantification using Quibit.

Selection of enzymes and fragment size range

We used DaCosta and Serenson et al. (2014) python script (*Digital_RADs.py*; available at https://github.com/BU-RAD-seq) to simulate the digestion process using a reference genome. The script looks for enzyme cutting sites and counts the number of fragments generated for a previously established size range and return the amount fragments with cutting sites (start – end): enzyme 1 – enzyme 2, enzyme 2 – enzyme 1, enzyme 1 – enzyme 1, enzyme 2 – enzyme 2, total enzyme 1 and total and enzyme 2. Given the absence of a humpback whale genome, to simulate the cleavage process we used the unmasked genome of *Balaenoptera acutorostrata* (BalAcu1.0 available at ncbi.nlm.nih.gov). Different enzyme combinations and fragment size range were tested in order to generate a given quantity and size of ddRAD loci. We choose the combination of enzymes that according with the simulation described above would generate around 30,000 fragments in a size range between 278 - 458bp, which ideally would allow a covering >100x considering the sample size.

RADseq Library Preparation and size selection

We used the ddRAD-seq protocol described by DaCosta and Sorenson (2014) for the library preparation. This consists of a single digestion reaction with two restriction enzymes (RE). The digestion was prepared with SphI and EcoRI (New England Biolabs) restriction enzymes at a concentration of 20U/µl and incubated overnight (16h) in the thermocycler at a constant temperature of 37ºC. Following digestion, were made barcoded (P1) and index (P2) adapters ligation. As in the protocol the P1 and P2 adapters include amplification and sequencing primer sequences besides six base pair barcodes and four more nucleotides (CATG for P1 and AATT for P2) overhang that match with the sticky-end left by SphI and EcoRI, respectively. We used three different P1 and ten P2 for do different arrangements for each sample. The proportion of these adapters was calculated as suggested Petersen et al. (2012), and the final concentration adjusted to 50 pmol/µl.

Samples were pooled and run on twelve lanes in a 2% low-melt agarose gel. We added internal size standards of 400 - 580bp (278-458bp fragments + 122bp adapter) to each lane to accurate the size selection. After, we extracted a slice from the gel correspondent to the size range doing a tapered cut, gradually reducing the area of the smaller fragments to approximately half the area of the larger fragments, because in some circumstances smaller fragments could be preferentially amplified in PCR (Walsh et al. 1992; DaCosta and Sorenson 2014). The extraction was made following the manufacturer's protocol using PureLink® Quick Gel Extraction Kit. Then we used the Phusion High-Fidelity PCR Master Mix (Thermo Scientific™) for twenty PCR cycles and purified with AMPure SPRI Beads (Beckman Coulter, Inc.). The samples were pooled, concentrated using SpeedVac™ and eluted in 40µl ultrapure water. Pool concentration was estimated with quantitative PCR (qPCR) checking different dilutions using a KAPA Biosystems kit. In the last step a new pool was made adjusting the concentration and amount of samples. This library was sequenced in a single lane at The Center for Applied Genomics, Hospital for Sick

Childern, Canada, with Illumina sequencing technology on a Hiseq 2000 system generating single-end sequences with 101 base pairs.

Bioinformatics

The sequence reads in fastq format were previously separate by index identification. We used ten different indexes (P2) which means that our files had three samples each, identify by barcodes at the P1 adapter. To pass samples into individuals files demultiplex was necessary, which was done with the program process_radtags provided with software STACKS (Catchen et al. 2011, 2013). Process_radtags also allows setting some parameters that we used to drop low quality scores reads and rescue barcodes and RAD-tags (catalog loci in STACKS). Individual samples were checked with FastQC tool, design for high throughput sequence data, which assess the overall quality and quantity of data per sample. Based on this evaluation two samples which did not have good sequencing results were discarded.

The remaining samples were aligned to masked Minke Whale (*B. acutorostrata)* genome in Bowtie2 (Langmead and Salzberg 2012). For this step, we used the masked genome in order to avoid repetitive regions. After, we used STACKS software that provided some pipes options to analyze and filter reads. We used four of these, pstacks, cstacks, sstacks and populations respectively. On cstacks we set the parameter of mismatches allowed between sample tags (n=3) and in pstacks two parameters were set, minimum coverage (m=35) and SNP (single nucleotide polymorphism) calling model with upper bound for epsilon 0.05. At the end of these steps we use the VCFtools software (Danecek et al. 2011) with a vcf output format from populations pipe to check the generated catalog. Thus we saw that were three samples with much less data than others, so these samples were discarded and we repeated the analysis from the first STACKS pipe with the remaining 25 samples. The last pipe, populations, output the filtered results with different formats and options that were set according to the following analyses.

Fasta_strict and vcf outputs of STACKS were used to check the amount of loci and SNPs obtained after filter the data. Bedtools software (Quinlan et al. 2010) and vcfR (Knaus and Grünwald 2016) package for R were used to check the range coverage through samples and loci, and the SNP distribution.

Demographic history

ADMIXTURE 1.3 (Alexander et al. 2009) with default parameters and the final set of polymorphic loci was used to test a possible population structure in our sample.

A custom python script (available at https://github.com/mgharvey/misc_python) was used to convert the fasta_strict format output from *populations*, into individual locus fasta files. For our first analysis, we selected only loci that are present in at least 70% of the individuals. Bayesian approach were employed with Migrate-n v.3.6.11 (Beerli, 2006) to estimate the theta parameter ($4N_e\mu$, μ is the mutation rate per generation) and the demographic history. The prior of the theta parameter was set to an exponential distribution with mean 0.005, minimum of 0 and maximum 0.1, after some tests runs. The final run of a single long chain, with 10,000 recorded steps saved every 25 steps and the 20,000 samples discarded as burn-in. We also used the skyline plot approach in migrate-n to estimate the demographic history of the populations. To estimate the absolute Ne values from the theta values we have to use a mutation rate. Previous studies based on intron sequences suggested the mutation rate for humpback whales is $\mu = 4.4 \times 10^{-10}$ (95% CI: 3.66×10^{-10} - 5.29×10^{-10}) per site per year (Jackson et al. 2009; Ruegg et al. 2013). We used a generation time of 18 (Chittleborough 1965, Roman and Palumbi 2003). We converted the Ne estimates to census size (Nc) by multiplying the former by 3.6, which has been used in previous studies (Roman and Palumbi 2003; Alter et al. 2007, 2012; Ruegg et al. 2010, 2013).

To test different scenarios for the recent demographic history of Brazilian humpback whale related to the whaling period we used the Approximate Bayesian Computation (ABC) approach

(Beaumont et al. 2002). As this is a computationally intensive method, to reduce the number of loci we increased the percentage of loci that must be present per individual to 85%. We also maintained only the polymorphic loci. Summary statistics were calculated using the msABC program (Pavlidis 2010), and the conversion of individual locus fasta files to the msABC input was made with fas2ms.pl script provided. We tested four demographic scenarios for the whaling period (Figure 1, Table 1). The first one is a constant population scenario (no size change), scenario 2 is a population that has been expanding since the last 2-10 generations, scenario 3 is a population that has been reducing since the last 2-10 generations and scenario 4 consisted of a population that has suffered a bottleneck between 2 and 10 generations and has been expanding ever since. The ranges of the priors were based on available information about commercial whaling (see introduction) and results of test runs (10,000 simulations) with different ranges. The prior parameters were random variables drawn from uniform distributions. Final results were based on 10^6 simulations for each scenario. We used the mutation rate and generation times as described above.

Principal components analysis (PCA) was used to verify the summary statistics prior distribution using randomly taken 10^4 summary statistics simulations of each scenario. We used the prcomp function in R 3.3.1 (R Core Team 2016) and the scatterplot generated by the ggbiplot R package (available at https://github.com/vqv/ggbiplot) to visualize the differences between scenarios and to compare them with the observed summary statistics. Posterior probability was calculated for each scenario using the abc R package (Csilléry et al. 2012) in using two selection methods with four threshold tolerance values (10%, 5%, 1% and 0.1%). Cross validation for ABC was performed with the same package to evaluate the effect of the different tolerance rates on the quality of the estimations and choose the more adequate. The selection methods were the multinomial logistic regression method (‘mnlogistic’) and the neural network approach (‘neuralnet’). Posterior parameters were calculated for the model with the highest posterior probability using ‘abc’ function of the abc R package. Based on ‘neuralnet’ method we did post

rejection adjustments and we also did logit transformations within the prior parameter values (Blum and François 2010).

Results

Sequencing and Alignment

The individualization of samples and removal of barcodes with process_radtags pipe at STACKS resulted in a total of 76,717,317 reads with 95pb, ranging between 373,541 and 15,680,891 reads per sample. The two samples that did not have good sequencing results, with 373,541 and 522,070 reads, were removed, resulting in a minimum of 798,065 reads per sample and a total of 75,821,706 reads (Supplementary Table S1). Of these, an average of 15.18% did not align to Minke masked genome, 72.35% aligned exactly 1 time and 12.45% aligned more than once, resulting in an overall alignment rate of 84.81%.

Catalog and SNP-callig

An initial catalog in STACKS was generated with the aligned reads of the 28 samples, but three samples presented too much missing data (> 30%) and, therefore, they were discarded and a new catalog were generated with the remaining twenty-five samples. We obtained 5,145 loci through these samples based on minke whale reference genome, and a total of 2,249 SNPs in our *M. novaeangliae* sample. The average read coverage of the loci was of 107.78 reads (Figure 2) and the SNPs frequency on these ranged from 1 to 11 (Supplementary Figure S1). The SNPs coverage was similar through the samples ranging from 23 to 234 except for one sample that had a upper range coverage from 344 to 571 (Figure 3).

Demographic history

ADMIXTURE software results showed no indication of population structure, since cross-Validation error indicates K values minimum and maximum of 0.6 and 1.2 respectively (Figure 4).

We used migrate-n with the final set of 5,145 loci to estimate the current nucleotide diversity (theta parameter). Migrate-n estimate of theta per site was 0.00237 (95% CI = 0.00067 – 0.004), which translated to a long-term effective population size (Ne) of ~75,000 (CI = 21,000-126,000) and a census population size (Nc) point estimate as high as ~270,000 (CI=76,000-450,000). The skyline method suggests the population size was much higher (by more than an order of magnitude) in the distant past, pointing for a long-term size decline for hundreds of thousands of years (Figure 5).

For the ABC approach, we selected a reduced set of loci consisting of polymorphic loci presented in at least 85% of the individuals, which resulted in 984 loci with a total of 1,187 SNPs. The theta per site for this dataset was 0.00282. In our posterior probability analyses (Figure 6) the tolerance values lower than 1% was not supported for any of the two selection method tested. According to the cross validation test that evaluate the effect of the different tolerance values on estimation quality, the three used here (1%, 5%, 10%) have equal effect. The most supported scenario seems the constant population although the recent expansion is also supported in some situations (see Figure 6). However, the value estimated in the recent expansion scenario for the ratio between the present and ancestral populations converge to one, converging the expansion scenario in the constant size scenario (see supplementary figures 2 and 3), suggesting the constant population size scenario as the most supported by this data set. The posterior distribution of the Ne on the constant size scenario, using ‘neuralnet’ method with tolerance of 5%, resulted a mean 124,002 (95% CI = 115,765 – 131,556 individuals) (Supplementary Figure 3).

Discussion

In this study, we used the RADseq approach to sequence thousands of loci on 25 biopsies of the Southwestern Atlantic humpback whales, corroborating other studies that applied similar approaches in non-model species (e.g. Garvin et al. 2010, Maughan et al. 2009).

Here we did not find evidence of any substructure within this population, corroborating previous studies (Cypriano-Souza et al. 2010; in press), suggesting that this high diversity is not caused by substructure or any significant gene flow in recent times. Our data set showed a high nuclear diversity for this humpback population agreeing with other studies with other markers in the same population (Cypriano-Souza et al. 2010), and other southern humpback stocks (e.g. Garrigue et al. 2004, Olavarria et al. 2007, Pomilla and Rosenbaum 2006, Rosenbaum et al. 2009, Valsecchi et al. 2002). Although there is no similar study (RADseq) to compare our results, the nucleotide diversity per site (theta) of ~0.237% is higher than theta found in the genome of a fin whale (0.15%) and three minke whales (0.06%) (Yim et al. 2014). However, this value is comparable with our previous study that sequenced 39 introns in the same population that found a theta of 0.13% (Cypriano-Souza et al 2014). The nucleotide diversity from the RADseq results is higher than that found with intron sequences since likely intron regions have lower substitution rates than RAD loci that are mostly located on non-functional intergenic regions (not shown).

The theta estimated with 5,145 loci with migrate-n translates, using Ruegg et al. (2013) mutation rate, to a point estimate of the effective population size of ~75,000 and an Nc ~270,000. However, as pointed above it is likely that our RAD loci may have a higher mutation rate than the loci used by Ruegg et al. (2013): comparing the nucleotide diversity of the two sets of markers in the same population suggests our RAD loci mutation rate may be ~1.8 times higher than the intron rate used before. Using this rate adjustment, BSA Ne point estimate would be around ~40,000 and the Nc ~140,000 (CI = ~41,000 - 250,000). In addition, the Ne estimated using the ABC approach with the stationary scenario using the intron mutation rate was ~124,000. However, if we adjust the

mutation rate the same way we did above (by the rate between the theta estimated from the intron and the RAD loci used in the ABC analyses) the posterior distribution for the Ne of the BSA population is now 57,000 (CI = ~53,000 – 60,000) with a respective Nc around 200,000.

The above estimates of humpback whale abundance, in special the estimated using the 5,145 loci, are very similar to our previous estimate for this population (Cypriano-Souza2013). Previous genetic estimates of abundance of other breeding stocks of the humpback whale were also in general very high. For example, point estimates for the North Atlantic populations based on mtDNA suggested between ~250,000 and ~150,000 individuals (Roman and Palumbi 2003; Alter and Palumbi 2009) and for the nuclear intron data set ~110,000 individuals (Ruegg et al. 2013). Besides, previous estimates suggested that humpback whale abundance is higher in the Southern Hemisphere than in the Northern (e.g. Ruegg et al. 2013; Cypriano-Souza 2013). Finally, genetic abundance point estimates of other baleen whales based on nuclear introns also found very high number, such as ~100,000 for North Pacific Ocean gray whales (Alter et al. 2012) and as high as a 670,000 for the smaller size Antarctic minke whale (Ruegg et al. 2010).

However, although we did not find evidence of population substructure in our sample, we could not discard the effect of past gene flow of BSA with other breeding stocks, which would have increased its long-term genetic diversity. Actually, previous studies based on mtDNA shown that several Southern Hemisphere stocks, although retaining their identity, present low to very low genetic differentiation with evidence of limited gene flow, in especially from South Atlantic and Indian Ocean (Rosenbaum et al. 2009; Rosenbaum et al. in press). In addition, the skyline plot result suggest an increasing population size for hundreds of thousands of years in the past (Fig. 5), that most likely extend to before the divergence of most Southern Hemisphere breeding stocks or even before the diversification of most humpback whale lineages (Jackson et al. 2009). Considering these evidences, it is likely that most of the increased population size in the distant past observed here represents the size of a metapopulation over the whole Southern Hemisphere or even the whole

species. However, this hypothesis should be tested with the addition of other populations and appropriate demographic scenarios.

Concerning the intense anthropogenic population bottleneck during the industrial whaling period, we could not detect any significant evidence of a genetic bottleneck using our data with the ABC approach (Figure 6). The intensity of the bottleneck (and therefore the likelihood to leave a significant signal) is a product of the number of generations and the absolute population size during the bottleneck (Amos 1996; Frankham et al. 2002). Given its large generation time and longevity, that the period of very reduced population size last at most only a few generation and the minimum size was around 500 individuals, it is not unexpected the absence of significant signal of an anthropogenic bottleneck. This is in agreement with previous studies on this population with mtDNA and microsatellite data (Cypriano-Souza et al. 2010, in preparation, Engel et al. 2008). On the other hand, Alter et al. 2012 in their study using mitochondrial control region sequences and an ABC approach detected a recent bottleneck (approximately 6 generations) in eastern Pacific gray whales. Although this species have similar generation time than the humpback whale they included ancient samples from pre-commercial whaling in the analyses that may have allowed them to detect this significant signal.

Funding

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [to J.D.E] and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [to S.L.B.].

Acknowledgements

We would like to thank the whole research team of the Humpback Whale Institute for the biopsy and collection samples supplied. We also thank the colleagues of lab of Genomic and

Molecular Biology (PUCRS) for their help in lab activities in special Fernanda Pedone for laboratory assistance.

References

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19:1655–1664

Alter SE and 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. Pre-Whaling Genetic Diversity and Population Ecology in Eastern Pacific Gray Whales: Insights from Ancient DNA and Stable Isotopes. PLoS ONE, 7(5):e35039

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

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

Beaumont M, Zhang W, Balding D. 2002. Approximate Bayesian Computation in Population Genetics. Genetics, 162:2025–2035

Beerli P. 2006. Comparison of Bayesian and maximum likelihood inference of population genetic parameters. Bioinformatics, 22(3):341–345

Best PB. 1994. A review of the catch statistics for modern whaling in southern Africa, 1908-1930. Rep int Whal Commn, 44:467-85

Blum MGB and François O. 2010.Non-linear regression models for Approximate Bayesian Computation. Stat Comput, 20: 63

Brodie, PF. 1975. Cetacean energetic: an overview of intraspecific size variation. Ecology, 56:152-161

Catchen JM, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes Genomes Genetics, 1:171-182

Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. Molecular Ecology, 22:3124-3140

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

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

Corkeron PJ and Connor RC. 1999. Why do baleen whales migrate?. Marine Mammal Science, 15(4): 1228-1245

Csilléry K, François O, Blum MGB. 2012. abc: an R package for approximate Bayesian computation (ABC). Methods in Ecology and Evolution, 3:475-479

Cypriano-Souza AL, Engel MH, Caballero S, Olavarría C, Flórez-González L, Capella J, Steel D, Sremba A, Aguayo A, Thiele D, Baker CS, Bonatto SL. 2010. Genetic differentiation

between humpback whales (Megaptera novaeangliae) from Atlantic and Pacific breeding grounds of South America. Marine Mammal Science

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

Cypriano-Souza AL. 2013. Estrutura populacional e história demográfica das populações de baleias jubarte (Megaptera novaeangliae) da América do Sul. Tese (Doutorado em Zoologia, programa de pós-graduação em biociências- zoologia) – Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul

Cypriano-Souza1 AL, Silva1TF, Engel MH, Bonatto SL. 2014. Genetic diversity and demography of Southwestern Atlantic humpback whale (Megaptera novaeangliae) population (Breeding Stock A). Unpublished paper submited to J Hered

DaCosta JM and Sorenson MD. 2014. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS ONE, 9(9): e106713

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27(15): 2156–2158

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

Engel M. 1996. Comportamento reprodutivo da baleia jubarte (Megaptera novaeangliae) em Abrolhos. Em C. Ades (Org.), Anais do XIV Encontro Anual de Etologia. Uberlândia, MG: Sociedade Brasileira de Etologia, pp 275-284

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

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

Gambell R. 1973. How whales survive. In: N. Calder (ed.) Nature in the Round. Weidenfeld & Nicolson, London, pp 193-202

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

Garvin MR, Saitoh K, Gharrett AJ. 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. Molecular Ecology Resources, 10:915-934

IUCN. International Union for the Conservation of Nature and Natural Resources - red list of threatened species. Available at http < http://www.iucnredlist.org/>

IWC. International Whaling Commission. Avaiable at http: < http://iwc.int>

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-González L, Palumbi SR. 2009. Big and slow: phylogenetic estimates of molecular evolution in baleen whales (suborder mysticeti). Mol Biol Evol 26(11):2427-40

Juliao HP. 2013. Abundancia e distribuição da baleia jubarte (Megaptera novaeangliae) na costa do Brasil. Master´s dissertation, Rio Grande (RS), Brazil: FURG, 128pp

Knaus BJ and Grünwald NJ.2016. VcfR: an R package to manipulate and visualize VCF format data. BioRxiv, 041277

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

Langmead B and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods, 9:357–359

Maughan PJ, Yourstone SM, Jellen EN, Udall JA. 2009. SNP discovery via genomic reduction, barcoding, and 454-pyrosequencing in amaranth. Plant Genome, 2:260-270

Olavarria C, Baker CS, Garrigue C, Poole M, Hauser N, Caballero S, Florez-Gonzalez 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

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

Pavlidis P, Laurent S, Stephan W.2010. msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis.Molecular Ecology Resources, 10:723-727

Payne R. 1995. Among whales. Scribner, New York, NY, 432pp

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS ONE, 7:e37135

Pomilla C and 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

Quinlan AR and Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 6, pp 841–842

Roman J and 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, Meyer 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

Tonnessen JN and Johnsen AO. 1982. The History of Modern Whaling. C. Hurst & Co., London, i-xx+798pp

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

Wade PR and Perryman W. 2002. An assessment of the eastern gray whale population in 2002. Unpubl. doc. submitted to Int. Whal. Comm. (SC/54/BRG7) 16 pp

Walsh PS, Erlich HA, Higuchi R. 1992. Preferential PCR amplification of alleles: mechanisms and solutions. Genome Research, 1:241–250

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

Williamson, GR. 1975. Minke whales off Brazil. Sci Rep Whales Res Inst, Tokyo, 27:37-59

Yim HS, Cho YS, Guang X, Kang SG, Jeong JY, Cha SS, Oh HM, Lee JH, Yang EC, Kwon KK, Kim YJ, Kim TW, Kim W, Jeon JH, Kim SJ, Choi DH6, Jho S, Kim HM, Ko J, Kim H, Shin YA, Jung HJ, Zheng Y, Wang Z, Chen Y, Chen M, Jiang A, Li E, Zhang S, Hou H, Kim TH, Yu L, Liu S, Ahn K, Cooper J, Park SG, Hong CP, Jin W, Kim HS, Park C, Lee K, Chun S, Morin PA, O'Brien SJ, Lee H, Kimura J, Moon DY, Manica A, Edwards J, Kim BC, Kim S, Wang J, Bhak J, Lee HS, Lee JH. 2014. Minke whale genome and aquatic adaptation in cetaceans. Nat Genet 46(1):88-92

Zerbini AN, Ward E, Engel MH, Andriolo A, Kinas PG. 2006. 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

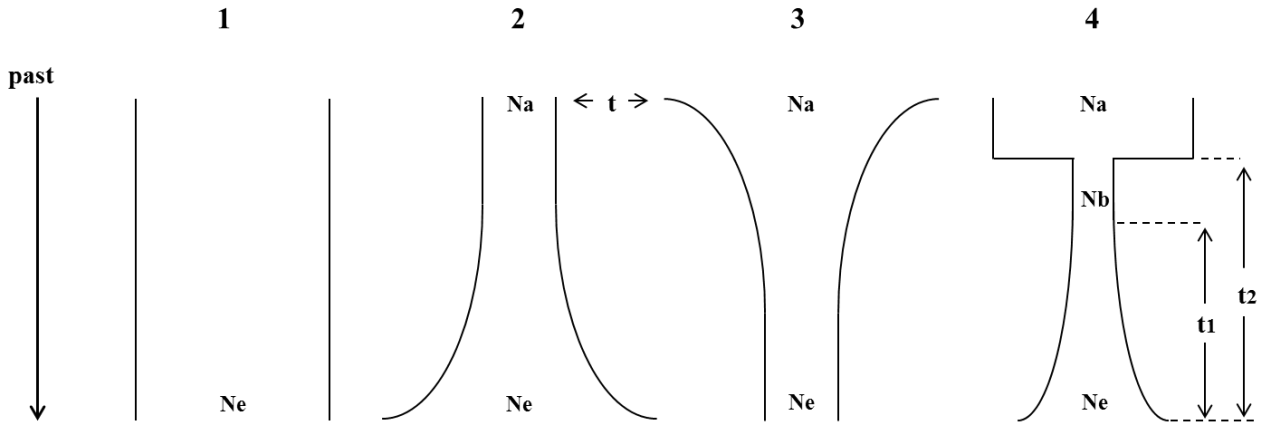


Figure 1. Demographic scenarios tested with the ABC approach: (1) constant population, (2) expanding population, (3) shrinking population and (4) bottleneck followed by expanding population. Parameters: effective size (Ne), ancestral size (Na), bottleneck size (Nb), time (t, t1, t2). Prior parameter distribution was in table 1.

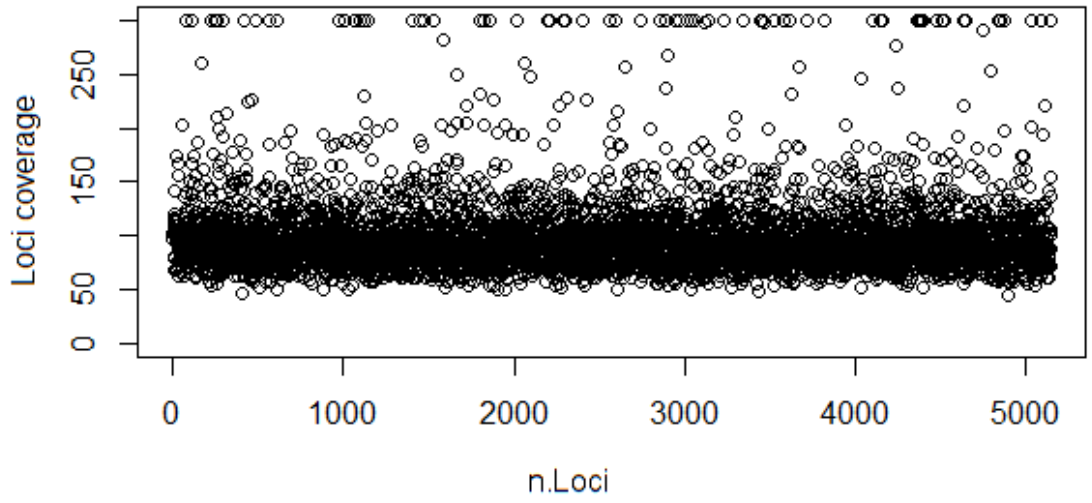


Figure 2. Sequencing coverage for each ddRAD loci. Coverage over 300 (1.38% of data) not appear on graphic to better visualization.

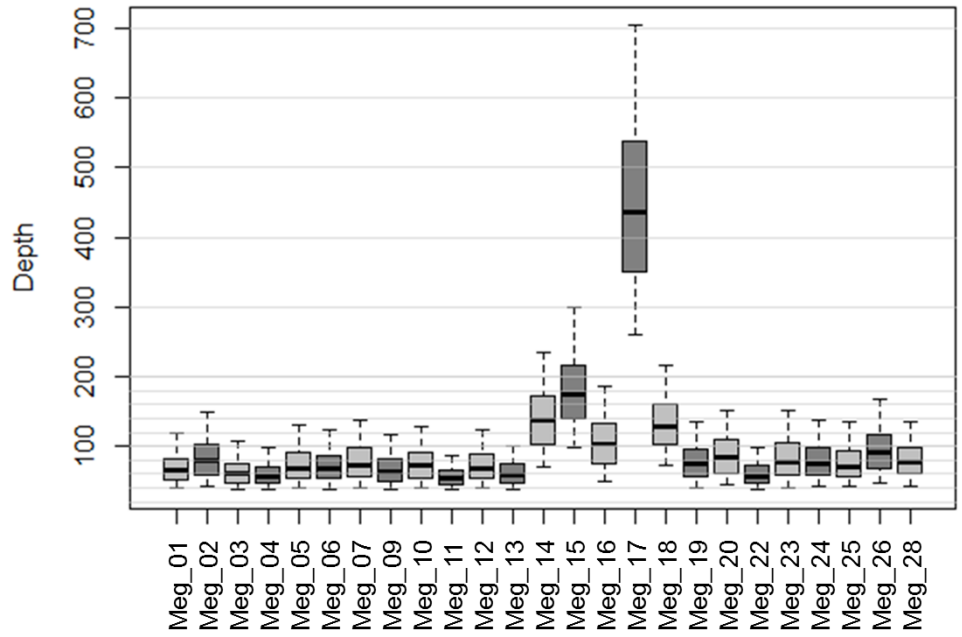


Figure 3. SNP coverage per sample.

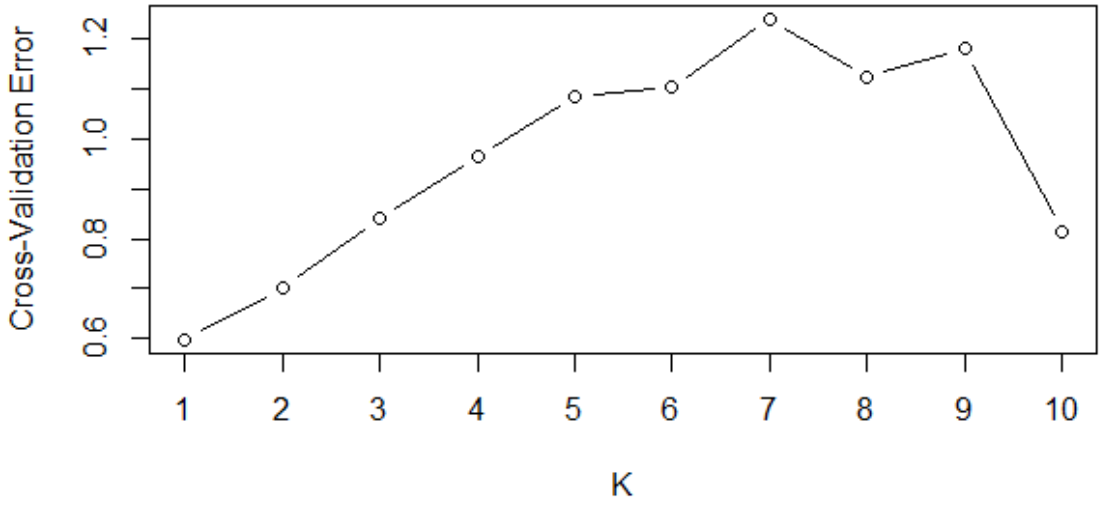


Figure 4. K values for Cross-Validation Error in Admixture software.

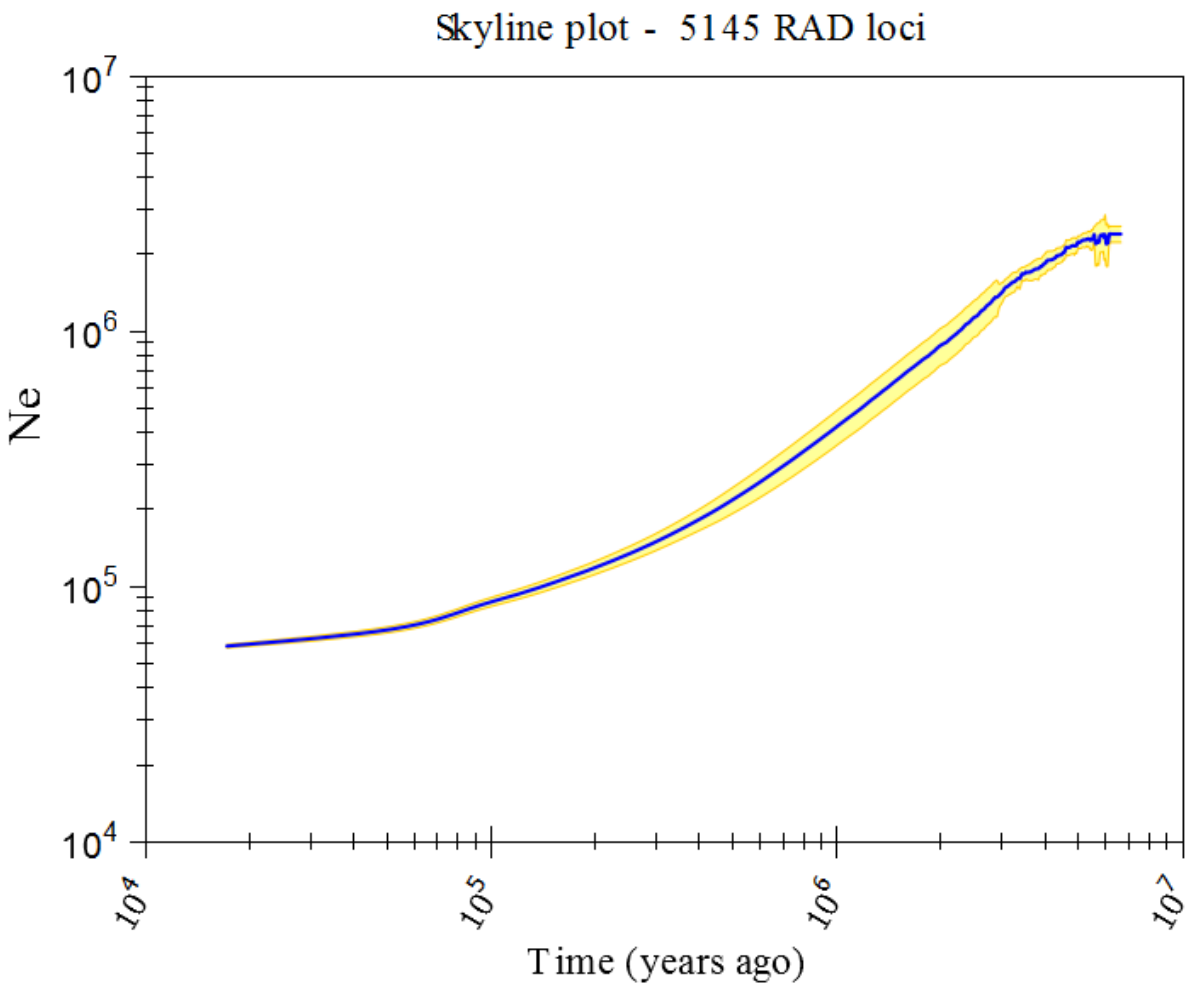


Figure 5. Skyline plot depicting the effective population size fluctuation over time estimated with migrate-n with 5145 loci. Blue solid line represents the median estimates and the yellow area denotes the 95% highest posterior densities (HPDs) for the estimates. Both axis are in the log scale.

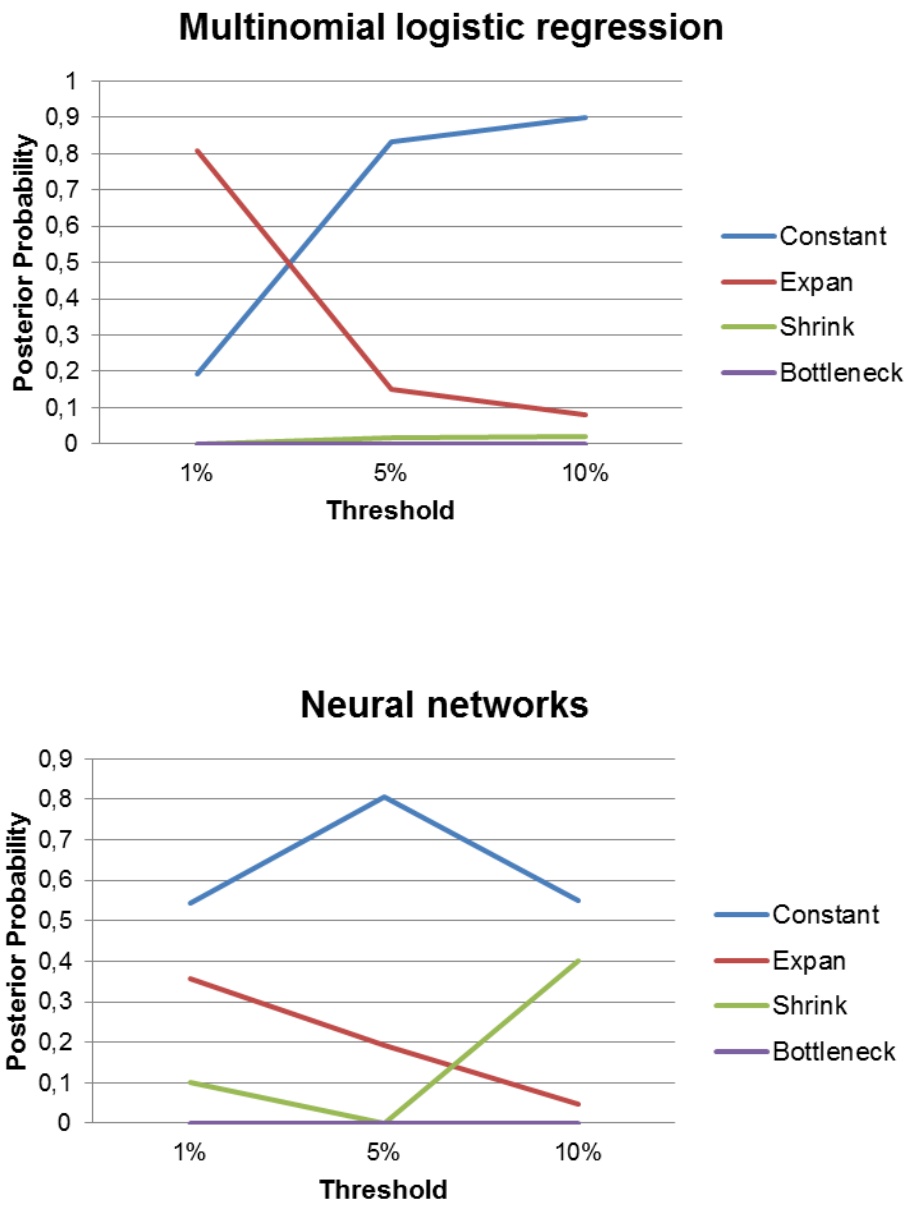


Figure 6. Posterior probabilities for the four scenarios in Multinomial logistic regression and Neural networks approaches with three threshold values.

Table 1. Priors for the parameters of the demographic scenarios (Fig. 1) used in msABC. Effective sizes (Ne) are in number of diploid individuals, ancestral size (Na) and bottleneck size (Nb) are proportions of the Ne. Times (t, t1 and t2) are in number of generations (generation time of 18 years). All used the uniform distribution.

Scenario parameter	**Minimum**	**Maximum**
Scenario 1 - constant		
Ne	80000	200000
Scenario 2 - expansion		
Ne	80000	200000
Na	0.01	1
t	2	10
Scenario 3 - shrink		
Ne	80000	200000
Na	2	10
t	2	10
Scenario 4 - bottleneck		
Ne	80000	200000
Nb	0.05	0.2
Na	2	10
t1	2	10
t2	t1+1	11

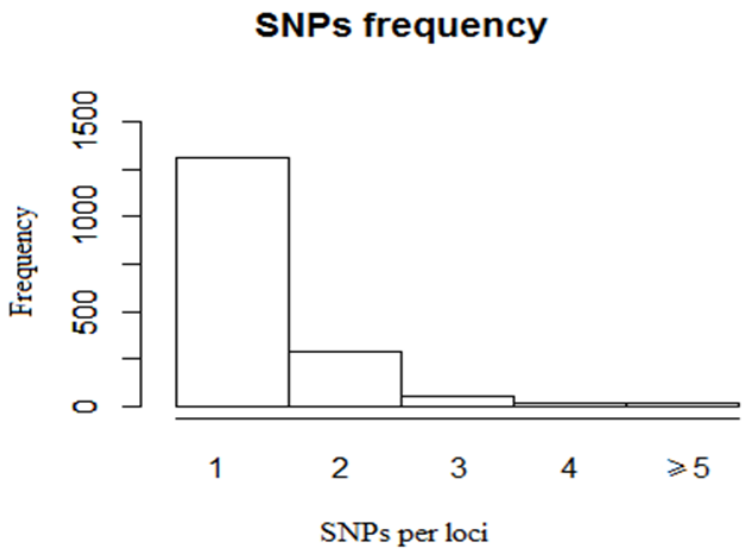
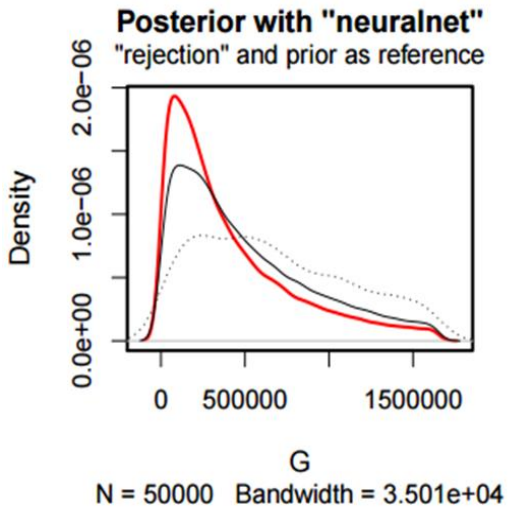
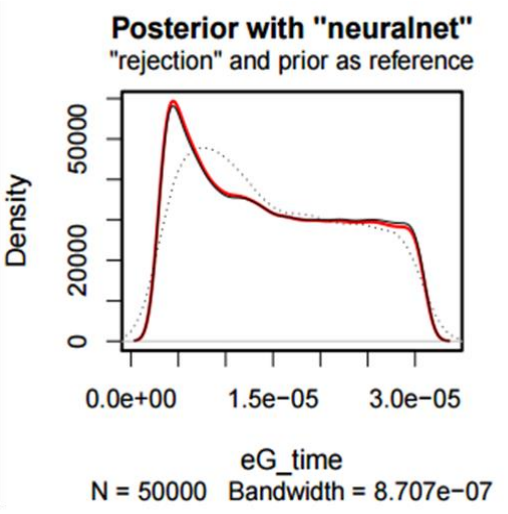


Figure Supplementary S1. SNPs frequency per loci. A total of 3,456 loci have no SNPs.

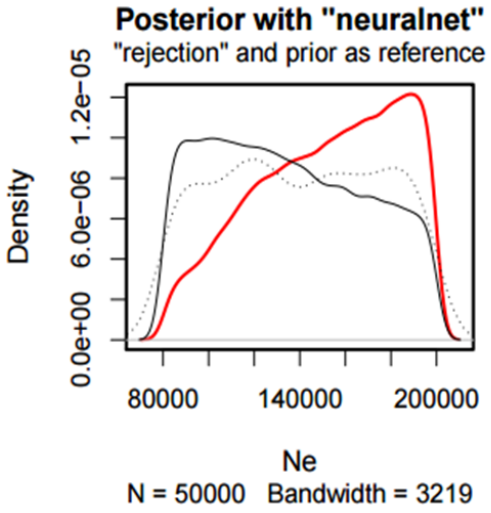
1



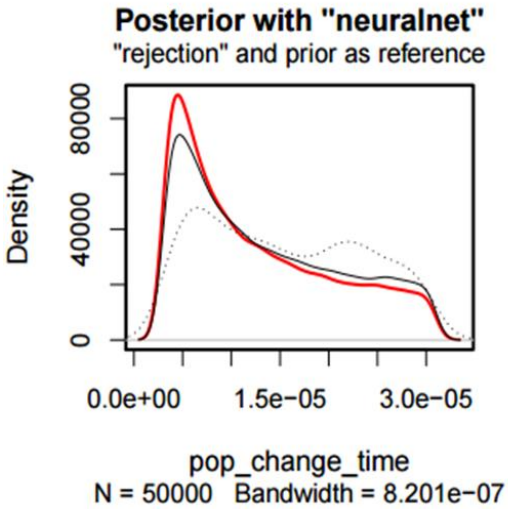
2



3



4



5

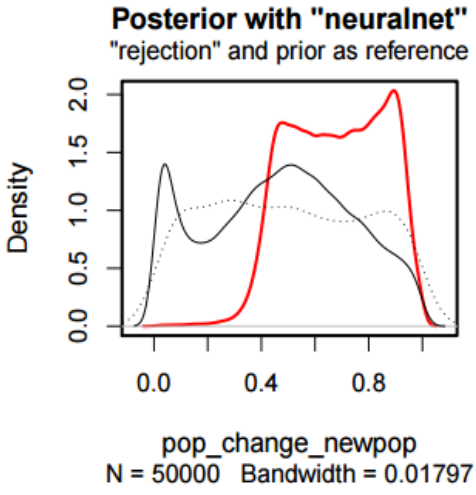


Figure Supplementary S2. Result of estimations for prior parameters: (1) Growth rate, (2) Change of the growth rate in the past, (3) Effective size, (4) Time the population start to expand and (5) Size ratio in relation to present size. Rates are given by $N(t) = N0 \exp^{-\alpha t}$, t is time before the present and N0 is present population size. All times are measures in units of 4N0 generation.

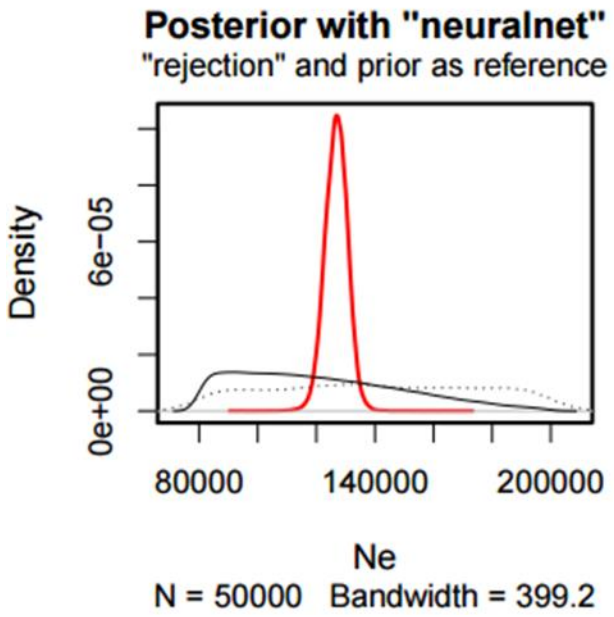


Figure Supplementary S3. Result of estimations for prior parameter Ne (current effective size) to constant population scenario using ‘neuralnet’ method with tolerance of 0.05 (5%).

Supplementary Table S1. Number of sequenced, discarded (not aligned) and retained reads after filtering by proccess_radtags.pl.

	Meg_01	%	Meg_02	%	Meg_03	%	Meg_04	%	Meg_05	%	Meg_06	%	Meg_07	%
Total Reads	2093924		2684033		1385448		1386188		2161040		2229732		2586464	
Not Aligned	344199	16.44%	431364	16.07%	238075	17.18%	236008	17.03%	339997	15.73%	325213	14.59%	420250	16.25%
Aligned 1 Time	1506423	71.94%	1901604	70.85%	977892	70.58%	992482	71.60%	1564402	72.39%	1627099	72.97%	1847416	71.43%
Aligned > 1 Time	243302	11.62%	351065	13.08%	169481	12.23%	157698	11.38%	256641	11.88%	277420	12.44%	318798	12.33%
Retained Reads	1749725	83.56%	2252669	83.93%	1147373	82.82%	1150180	82.97%	1821043	84.27%	1904519	85.41%	2166214	83.75%

	Meg_08	%	Meg_09	%	Meg_10	%	Meg_11	%	Meg_12	%	Meg_13	%	Meg_14	%
Total Reads	1178420		1856265		2649605		1199843		2395243		1294876		798065	
Not Aligned	391967	33.26%	289208	15.58%	438528	16.55%	173198	14.44%	394145	16.46%	185392	14.32%	124477	15.60%
Aligned 1 Time	671454	56.98%	1346585	72.54%	1891183	71.38%	882943	73.59%	1661585	69.37%	945723	73.04%	574473	71.98%
Aligned > 1 Time	114999	9.76%	220472	11.88%	319894	12.07%	143702	11.98%	339513	14.17%	163761	12.65%	99115	12.42%
Retained Reads	786453	66.74%	1567057	84.42%	2211077	83.45%	1026645	85.56%	2001098	83.54%	1109484	85.68%	673588	84.40%

Cont. Supplementary Table S1.

	Meg_15	%	Meg_16	%	Meg_17	%	Meg_18	%	Meg_19	%	Meg_20	%	Meg_21	%
Total Reads	5503950		7275458		4401163		15680891		5342512		2312942		2973542	
Not Aligned	903280	16.41%	1149023	15.79%	628927	14.29%	1952679	12.45%	789481	14.78%	310267	13.41%	402509	13.54%
Aligned 1 Time	3938897	71.56%	5168756	71.04%	3125697	71.02%	11900195	75.89%	3884300	72.71%	1729356	74.77%	2178572	73.27%
Aligned > 1 Time	661773	12.02%	957679	13.16%	646539	14.69%	1828017	11.66%	668731	12.52%	273319	11.82%	392461	13.20%
Retained Reads	4600670	83.59%	6126435	84.21%	3772236	85.71%	13728212	87.55%	4553031	85.22%	2002675	86.59%	2571033	86.46%

	Meg_22	%	Meg_23	%	Meg_24	%	Meg_25	%	Meg_26	%	Meg_27	%	Meg_28	%
Total Reads	1271512		1517861		2896955		2659506		2311585		3236693		3166986	
Not Aligned	196058	15.42%	257943	16.99%	486984	16.81%	440800	16.57%	316910	13.71%	463309	14.31%	498430	15.74%
Aligned 1 Time	924573	72.71%	1067744	70.35%	2051671	70.82%	1852631	69.66%	1692676	73.23%	2362405	72.99%	2283668	72.11%
Aligned > 1 Time	150881	11.87%	192174	12.66%	358300	12.37%	366075	13.76%	301999	13.06%	410979	12.70%	384888	12.15%
Retained Reads	1075454	84.58%	1259918	83.01%	2409971	83.19%	2218706	83.43%	1994675	86.29%	2773384	85.69%	2668556	84.26%

Referências Bibliográficas

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

Cypriano-Souza AL. 2013. Estrutura populacional e história demográfica das populações de baleias jubarte (*Megaptera novaeangliae*) da América do Sul. Tese (Doutorado em Zoologia, programa de pós-graduação em biociências- zoologia) – Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul

Cypriano-Souza1 AL, Silva1TF, Engel MH, Bonatto SL. 2014. Genetic diversity and demography of Southwestern Atlantic humpback whale (*Megaptera novaeangliae*) population (Breeding Stock A). Unpublished paper submited to J Hered

DaCosta JM and Sorenson MD. 2014. *Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol*. PLoS ONE, 9(9):e106713

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

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

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

Garvin MR, Saitoh K, Gharrett AJ. 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. Molecular Ecology Resources, 10:915–934

IWC. International Whaling Commission. Disponível em http: < http://iwc.int>

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

Johnson H and Wolman AA. 1984. The Humpback Whale, *Megaptera novaeangliae*. Marine Fisheries Review, 46(4):30-37

Maughan PJ, Yourstone SM, Jellen EN, Udall JA. 2009. SNP discovery via genomic reduction, barcoding and 454-pyrosequencing in amaranth. Plant Gen, 2:260-270

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

Tonnessen JN and Johnsen AO. 1982. The History of Modern Whaling. C. Hurst & Co., London, i-xx+798pp

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