TITLE:

Population history of Antarctic and common minke whales inferred from individual whole-genome sequences

AUTHOR(S):

Kishida, Takushi

CITATION:

Kishida, Takushi. Population history of Antarctic and common minke whales inferred from individual whole-genome sequences. *Marine Mammal Science* 2017, 33(2): 645-652

ISSUE DATE:

2017-04

URL:

<http://hdl.handle.net/2433/228168>

RIGHT:

This is the accepted version of the following article: [Takushi Kishida. Population history of Antarctic and common minke whales inferred from individual whole-genome sequences. *Marine Mammal Science*, 33(2): 645–652 (April 2017)], which has been published in final form at <https://doi.org/10.1111/mms.12369>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.; The full-text file will be made open to the public on 13 APR 2018 in accordance with publisher's 'Terms and Conditions for Self-Archiving'; This is not the published version. Please cite only the published version.; この論文は出版社版でありません。引用の際には出版社版をご確認ご利用ください。

- 1 Population history of Antarctic and common minke whales inferred from individual whole-
- 2 genome sequences
- 3
- 4
- 5 KISHIDA, Takushi
- 6 Wildlife Research Center, Kyoto University
- 7 2-24 Tanaka Sekiden-cho, Sakyo, Kyoto 606-8203, Japan
- 8 e-mail: taku.kishida@gmail.com
- 9
- 10
- 11
- 12
- 13
- 14 Keywords
- 15 *Balaenoptera bonaerensis*
- 16 *Balaenoptera acutorostrata*
- 17 PSMC
- 18
- 19

20 Speciation in the open ocean has long been studied (*e.g.*, Palumbi 1992, Miya and Nishida
21 1997, Davis *et al.* 2014, Friesen 2015), but it remains largely elusive how populations of highly
22 mobile animals, such as whales, in such an open environment become reproductively isolated.
23 Baleen whales of the genus *Balaenoptera* undertake extensive migrations, and there are few
24 obvious barriers that potentially isolate their populations in the open ocean. Intra-species
25 genetic diversities are extremely low in several species of these whales (Hoelzel 1994).

26 Among whales of the genus *Balaenoptera*, two species of minke whales, common minke whale
27 *Balaenoptera acutorostrata* and Antarctic minke whale *B. bonaerensis*, are closely related
28 (Pastene *et al.* 2007). Common minke whales are distributed worldwide, with subspecies
29 identified in the North Atlantic (*B. acutorostrata acutorostrata*), North Pacific (*B. acutorostrata*
30 *scammoni*), and Southern Ocean (dwarf minke whale). On the other hand, Antarctic minke
31 whales are distributed only in the Southern Hemisphere. Minke whales are considered to have
32 originated in the Southern Hemisphere, and common minke whales radiated into the North
33 Atlantic and the North Pacific after speciation (Pastene *et al.* 2007). Based on the
34 mitochondrial sequences, these two minke whale species are estimated to have diverged
35 approximately 7.5 million years ago (MYA) (Park *et al.* 2015) to 4.7 MYA in the Southern
36 Hemisphere (Pastene *et al.* 2007), and *B. a. acutorostrata* and *B. a. scammoni* diverged
37 approximately 2.4 MYA (Park *et al.* 2015) to 1.5 MYA (Pastene *et al.* 2007). Pastene *et al.*
38 (2007) argued that elevated ocean temperatures due to global warming in the Pliocene
39 facilitated allopatric speciation into two minke whale species by disrupting the continuous belt
40 of upwelling maintained by the Antarctic Circumpolar Current (ACC) in the Southern
41 Hemisphere. They also speculated that an overall increase in carrying capacity as ACC-driven
42 upwelling was re-established due to cooler temperature after the end of the Pliocene
43 facilitated population expansion in both minke whale species.

44 The distribution of the time since the most recent common ancestor (TMRCA) between alleles
45 in an individual gives information about the historical changes in effective population size (N_e)
46 over time. Based on this, a revolutionary model for inferring demographic history from whole-
47 genome sequence of a diploid individual, named pairwise sequential Markovian coalescent
48 (PSMC), was proposed (Li and Durbin 2011). In this model, numerous independent loci
49 contained in a diploid genome, each with its own TMRCA between the two alleles carried by an
50 individual, are analyzed. When the PSMC inferences of two closely-related species that share a
51 recent common ancestor are plotted together, the time point where the two historical
52 effective population sizes diverge approximates the divergence time because these two
53 species share the same N_e value before divergence, though there are several caveats to be
54 taken into account. For example, it is possible that the two populations had the same size after
55 the split, which will lead to an underestimate (Prado-Martinez *et al.* 2013).

56 Recently, whole genome shotgun (WGS) reads of a North Pacific common minke whale
57 sampled near the east coast of Korea were reported (Yim *et al.* 2014). In addition, my group
58 sequenced WGS of an Antarctic minke whale from the Antarctic Ocean (Kishida *et al.* 2015).
59 Using these data, historical N_e of common and Antarctic minke whales were inferred with
60 PSMC in order to further our understanding of demographic processes leading to the
61 differentiation of these two species. Because all common minke whale subspecies belonged to
62 the same ancestral population (*i.e.*, same N_e) before divergence, the North Pacific sample of *B.*
63 *a. scammoni* can be used to represent common minke whales for this comparison. Paired-end
64 short reads of these whales listed in Table 1 were aligned to the BalAcu1.0 minke whale
65 genome (Yim *et al.* 2014) with BWA (Li and Durbin 2009) after removal of adapter sequences
66 and low quality bases using Trimmomatic ver. 0.33 (Bolger *et al.* 2014) (Trimmomatic settings:
67 ILLUMINACLIP:\$HOME/Trimmomatic-0.33/adapters/TruSeq3-PE-2.fa:2:30:10 LEADING:20

TRAILING:20 SLIDINGWINDOW:5:25 CROP:100 MINLEN:36) . Duplicated reads were removed and scaffolds identified as putative sex chromosome regions by Yim *et al.* (2014, see their Supplementary Table 16) were excluded from further analyses. Methods of Li and Durbin (2011) were followed to obtain the diploid consensus sequences, which excluded the following sites: (1) the read depth is more than twice or less than one-third of the average depth, (2) the site is within 5 bp around a predicted short insertion or deletion, and (3) the root-mean-square mapping quality is low (<10) (Table 1). Based on these consensus sequences, the PSMC outputs were obtained with 64 atomic time intervals and 28 free interval parameters (the ‘-p’ parameter used for PSMC was set to ‘4+25*2+4+6’). Bootstrapping was performed 100 times by breaking the consensus sequences into 5Mb segments and randomly resampling a set of segments by replacement.

The PSMC plots thus obtained are shown in Fig. 1. Average mutation rate per base per year (μ) and average generation time (g) are required for scaling results of PSMC, but it is generally not easy to estimate these values. These values may differ across species and might have been changed through evolution. Currently, the generation time g is estimated to be approximately 22 yr in both species of minke whales (Taylor *et al.* 2007). Regarding mutation rate, analyzing only limited numbers of intronic loci, it was previously approximated that $\mu=5\times10^{-10}$ in baleen whales (Alter *et al.* 2007, Jackson *et al.* 2009). However recently, based on the genome-wide analyses, Yim *et al.* (2014) estimated that $\mu=1.07\times10^{-9}$ in common minke whales. In addition, it is estimated that $\mu=1.22\times10^{-9}$ in Yangtze River dolphins and $\mu=0.84\times10^{-9}$ in common bottlenose dolphins (Zhou *et al.* 2013). Therefore, I assumed that μ is approximately 1×10^{-9} in whales. Based on these assumptions ($g=22$, $\mu=1\times10^{-9}$), historical N_e was estimated for both species (Fig. 1).

Yim *et al.* (2014) also estimated the demographic history of North Pacific common minke whales with PSMC and obtained essentially the same results. In addition to the historical N_e of common minke whales, I estimated that of Antarctic minke whales and plotted both together in Fig. 1.

Based on the PSMC results, I found that the effective population size of *B. a. scammoni* decreased profoundly about 1.5-3 MYA, when common minke whales are thought to have immigrated to the Northern Hemisphere and diverged into two subspecies (Pastene *et al.* 2007). Over the same period of time, the effective population size of Antarctic minke whales was also reduced. These findings may suggest that there had been environmental deterioration for minke whales around the Antarctic at that time. However, it is also possible that these inferred decrease of N_e are caused by population structuring (Mazet *et al.* 2016), especially by population divergence (Foote *et al.* 2016).

Unfortunately, probably due to lower heterozygosity in the North Pacific minke whale genome (Yim *et al.* 2014), N_e of common minke whales is not plotted where $2\mu T > 6 \times 10^{-3}$ (before 3 MYA). This means that I could not find the exact point where *B. bonaerensis* and *B. acutorostrata* speciated, which is estimated to be approximately 4-5 MYA (Pastene *et al.* 2007). Prado-Martinez *et al.* (2013) proposed another approach to infer divergence time with PSMC. In this approach, they hybridized two haploid sequences from each species and then ran PSMC on the pseudo-diploid sequence thus obtained, and suggested that the time point where the inferred N_e goes to infinity corresponds to the divergence time. Following their methods, I generated the pseudo-diploid sequence of *B. bonaerensis* and *B. acutorostrata scammoni* using seqtk program provided by Heng Li (<https://github.com/lh3/seqtk>), and estimated the ancestral N_e with 5, 18, and 28 free interval parameters (Fig. 2). To derive haploid sequences from each minke whale species, an allele with low estimated consensus quality (<20) was excluded and

an allele at a heterozygous site was selected randomly (*i.e.*, the ‘-rq20’ option was used for seqtk). As a result, there were sequences of 1.43 Gbp in total remained for PSMC calculation.

Fig. 2 shows that PSMC infers excessively large effective population size at around $2\mu T = 2 \times 10^{-3} - 3 \times 10^{-3}$ (1-1.5 MYA, when $\mu = 1 \times 10^{-9}$) except in the case of 5 intervals. This is inconsistent with the previous report that *B. bonaerensis* and *B. acutorostrata* diverged some 4-5 MYA (Pastene *et al.* 2007), even if the mutation rate is smaller than my assumption (*e.g.*, 2-3 MYA, when $\mu = 5 \times 10^{-10}$, as suggested by Alter *et al.* 2007 and Jackson *et al.* 2009). Following these results, there are two hypotheses to be taken into account. First, two species of minke whales diverged at approximately $2\mu T = 3 \times 10^{-3}$ (1.5 MYA), much more recently than the previous estimations. Indeed, PSMC plots of Antarctic and common minke whale shown in Fig. 1 seem to diverge at around $2\mu T = 3 \times 10^{-3}$. And second, it is because there had been gene flow between these two minke whale species after speciation. Pastene *et al.* (2007) speculated that both *B. bonaerensis* and *B. acutorostrata* were distributed in the Southern Hemisphere before the divergence of *B. a. acutorostrata* and *B. a. scammoni* at around 1.5 MYA, and this inconsistency in the estimated divergence time of two minke whale species may be caused if there had been inter-species gene flow between populations of *B. bonaerensis* and *B. acutorostrata* when they lived sympatrically in the Southern Hemisphere (Fig. 3A). Following the second hypothesis, a model of population divergence mimicking the evolution of Antarctic and common minke whales with no changes in the total size of N_e is considered (Fig. 3B). Based on this model, PSMC plots of populations 1 and 2, generated using ms-HOT program (Hellenthal and Stephens 2007) modified by Heng Li (msHOT-lite, <https://github.com/lh3/foreign/tree/master/msHOT-lite>), are shown in Figs 3C (gene flow is not assumed) and 3D (symmetric gene flow is assumed). In Fig. 3D, sizes of both populations 1 and 2 are inferred to be decreased gradually at around $2\mu T = 2 \times 10^{-3}$ to 5×10^{-3} (1-2.5 MYA). The

same tendency is also observed in Fig. 1. These data suggest that the second hypothesis is also possible, that is there had been gene flow between populations of Antarctic and common minke whales when they lived in the Southern Hemisphere together (4.7-1.5 MYA), and that the apparent decrease in the inferred effective population sizes at this point in evolution was mainly caused by population divergence. In any case, after that, N_e of Antarctic minke whales increased, while there was not a profound increase in the effective population size of North Pacific minke whales. Such population history should contribute to current census population sizes of these whales —there are approximately 500,000-700,000 Antarctic minke whales in the Southern Hemisphere, whereas there are less than 30,000 common minke whales in the North Pacific (Thomas *et al.* 2016). Further analyses, including North Atlantic common minke whales and dwarf minke whales, will reveal speciation and divergence of minke whale populations more clearly.

Acknowledgements

I am grateful to Dr. Kentaro Fukuta (National Institute of Genetics) and reviewers of this manuscript for helpful comments. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics. This study was financially supported in part by JSPS KAKENHI (grant nos. 24770075 and 15K07184).

161

162 Figure legends

163

164 Figure 1

165 Demographic history of minke whales inferred with pairwise sequential Markovian coalescent.

166 Bootstrap replicates are represented by light narrow lines. The x axis gives time, and the upper

167 x axis indicates scaling in years with the assumption of $\mu=1\times 10^{-9}$. The y axis gives the effective

168 population size, and scaling in numbers with the assumption of $\mu=1\times 10^{-9}$ and $g=22$ is shown in

169 right.

170 Abbreviations and parameters: MYA; million years ago, KYA; kilo years ago, T ; time before

171 present [year], N_e ; effective population size, μ ; mutation rate per base per year, g ; generation

172 time [year].

173

174 Figure 2

175 Pairwise sequential Markovian coalescent (PSMC) plots of a pseudo-diploid genome sequence

176 of *B. bonaerensis* and *B. acutorostrata scammoni*. The ‘-p’ parameter used for PSMC is set to

177 ‘20+4*5’, ‘6+17*2’ and ‘4+25*2+4+6’ for 5, 18, and 28 intervals, respectively.

178

179 Figure 3

180 A) A phylogenetic tree of minke whales taken from Pastene et al. (2007). Minke whale
181 subspecies which were not analyzed in this study are represented by light narrow lines.
182 Because both *B. bonaerensis* and *B. acutorostrata* distributed in the Southern Hemisphere
183 sympatrically approximately 4.7 to 1.5 MYA (Pastene *et al.* 2007), it is possible that there had
184 been gene flow between these two whale populations at that time.

185 B) A model of population divergence mimicking the evolution of Antarctic and common minke
186 whales with no changes in the size of populations, which in total remains 40,000. An ancestral
187 population of size $N_{A1}=40,000$ splits 214,000 generations ago into two descending populations
188 of size $N_1=N_{A2}=20,000$, and the population of size $N_{A2}=20,000$ splits 68,200 generations ago
189 into two descending populations of size $N_2=N_3=10,000$. Populations 1, 2, and 3 mimic the
190 populations of *B. bonaerensis*, *B. a. scammoni*, and *B. a. acutorostrata*, respectively (dwarf
191 minke whale population is not considered).

192 Based on this model, with the assumption of $\mu=1\times 10^{-9}$ and $g=22$, demographic history of
193 populations 1 and 2 were inferred with pairwise sequential Markovian coalescent (C, D). Gene
194 flow between populations was not assumed in (C), whereas it was assumed that there is a
195 period of symmetric gene flow (migration rate per generation $m=0.001$) after divergence until
196 the split of populations 2 and 3 (214,000-68,200 generations ago) in (D). Each data set consists
197 of 100 blocks of 10 Mb (total of 1 Gb) sampled from a single diploid individual, generated with
198 the following command lines: (C) population 1, msHOT-lite 2 100 -l -t 8800 -r 7040 10000000 -l
199 3 2 0 0.0 -n 1 2 -n 2 1 -n 3 1 -G 0.0 -ej 1.705 3 2 -en 1.705 1 2 -en 1.705 2 2 -ej 5.35 2 1 -en 5.35
200 1 4; population 2, msHOT-lite 2 100 -l -t 8800 -r 7040 10000000 -l 3 2 0 0.0 -n 1 1 -n 2 2 -n 3 1 -
201 G 0.0 -ej 1.705 3 1 -en 1.705 1 2 -en 1.705 2 2 -ej 5.35 2 1 -en 5.35 1 4, and (D) population 1,
202 msHOT-lite 2 100 -l -t 8800 -r 7040 10000000 -l 3 2 0 0.0 -n 1 2 -n 2 1 -n 3 1 -G 0.0 -m 1 2 0 -m 2
203 1 0 -ej 1.705 3 2 -em 1.705 1 2 40 -em 1.705 2 1 40 -en 1.705 1 2 -en 1.705 2 2 -ej 5.35 2 1 -em

204 5.35 1 2 0 -em 5.35 2 1 0 -en 5.35 1 4; population 2, msHOT-lite 2 100 -l -t 8800 -r 7040
205 10000000 -l 3 2 0 0.0 -n 1 1 -n 2 2 -n 3 1 -G 0.0 -m 1 2 0 -m 2 1 0 -ej 1.705 3 1 -em 1.705 1 2 40 -
206 em 1.705 2 1 40 -en 1.705 1 2 -en 1.705 2 2 -ej 5.35 2 1 -em 5.35 1 2 0 -em 5.35 2 1 0 -en 5.35 1
207 4.

208

209 Literature Cited

210 Alter, S. E., E. Rynes and S. R. Palumbi. 2007. DNA evidence for historic population size and
211 past ecosystem impacts of gray whales. *Proceedings of the National Academy of*
212 *Sciences* 104:15162-15167.

213 Bolger, A. M., M. Lohse and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina
214 sequence data. *Bioinformatics*:btu170.

215 Davis, M. P., N. I. Holcroft, E. O. Wiley, J. S. Sparks and W. Leo Smith. 2014. Species-specific
216 bioluminescence facilitates speciation in the deep sea. *Marine Biology* 161:1139-1148.

217 Foote, A. D., N. Vijay, M. C. Avila-Arcos, *et al.* 2016. Genome-culture coevolution promotes
218 rapid divergence of killer whale ecotypes. *Nature Communications* 7:11693.

219 Friesen, V. L. 2015. Speciation in seabirds: Why are there so many species...and why aren't
220 there more? *Journal of Ornithology* 156:27-39.

221 Hellenthal, G. and M. Stephens. 2007. msHOT: modifying Hudson's ms simulator to incorporate
222 crossover and gene conversion hotspots. *Bioinformatics* 23:520-521.

223 Hoelzel, A. R. 1994. Genetics and ecology of whales and dolphins. *Annual Review of Ecology*
224 *and Systematics* 25:377-399.

- 225 Jackson, J. A., C. S. Baker, M. Vant, D. J. Steel, L. Medrano-González and S. R. Palumbi. 2009.
- 226 Big and slow: Phylogenetic estimates of molecular evolution in baleen whales
- 227 (Suborder Mysticeti). *Molecular Biology and Evolution* 26:2427-2440.
- 228 Kishida, T., J. G. M. Thewissen, T. Hayakawa, H. Imai and K. Agata. 2015. Aquatic adaptation
- 229 and the evolution of smell and taste in whales. *Zoological Letters* 1:9.
- 230 Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler
- 231 transform. *Bioinformatics* 25:1754-1760.
- 232 Li, H. and R. Durbin. 2011. Inference of human population history from individual whole-
- 233 genome sequences. *Nature* 475:493-496.
- 234 Mazet, O., W. Rodriguez, S. Grusea, S. Boitard and L. Chikhi. 2016. On the importance of being
- 235 structured: instantaneous coalescence rates and human evolution—lessons for
- 236 ancestral population size inference? *Heredity* 116:362-371.
- 237 Miya, M. and M. Nishida. 1997. Speciation in the open ocean. *Nature* 389:803-804.
- 238 Palumbi, S. R. 1992. Marine speciation on a small planet. *Trends in Ecology & Evolution* 7:114-
- 239 118.
- 240 Park, J., Y. R. An, N. Kanda, *et al.* 2015. Cetaceans evolution: insights from the genome
- 241 sequences of common minke whales. *BMC Genomics* 16:13.
- 242 Pastene, L. A., M. Goto, N. Kanda, *et al.* 2007. Radiation and speciation of pelagic organisms
- 243 during periods of global warming: The case of the common minke whale, *Balaenoptera*
- 244 *acutorostrata*. *Molecular Ecology* 16:1481-1495.
- 245 Prado-Martinez, J., P. H. Sudmant, J. M. Kidd, *et al.* 2013. Great ape genetic diversity and
- 246 population history. *Nature* 499:471-475.

- 247 Taylor, B. L., S. J. Chivers, J. Larese and W. F. Perrin. 2007. Generation length and percent
248 mature estimates for IUCN assessments of cetaceans. Administrative report LJ-07-01,
249 Southwest Fisheries Science Center, National Marine Fisheries Service.
- 250 Thomas, P. O., R. R. Reeves and R. L. Brownell. 2016. Status of the world's baleen whales.
251 Marine Mammal Science 32:682-734.
- 252 Yim, H.-S., Y. S. Cho, X. Guang, *et al.* 2014. Minke whale genome and aquatic adaptation in
253 cetaceans. Nature Genetics 46:88-92.
- 254 Zhou, X., F. Sun, S. Xu, *et al.* 2013. Baiji genomes reveal low genetic variability and new insights
255 into secondary aquatic adaptations. Nature Communications 4:2708.

Table 1. Sequence data used for PSMC.

species	Sequence Read Archive accession nos.	average depth ¹	callable ²
<i>B. acutorostrata scammoni</i>	SRR893003, SRR896642, SRR901891	51.1	1.49Gbp
<i>B. bonaerensis</i>	DRR014695	82.2	1.45Gbp

¹ Average depth was calculated after removal of duplicated reads and sex chromosomal reads

² Total length of regions for calculation. Following sites are excluded: (1) the read depth is more than twice or less than one-third of the average depth, (2) the site is within 5 bp around a predicted short indel, and (3) the root-mean-square mapping quality is low (<10).