Received Date : 10-Dec-2014
Revised Date : 06-Mar-2016
Accepted Date : 22-Mar-2016
Article type : Original Article

Corresponding author mail id : alana.alexander@ku.edu

RESEARCH ARTICLE: Molecular Ecology

What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)?

Alana Alexander[1,2,3*], Debbie Steel[1,2], Kendra Hoekzema[2], Sarah Mesnick[4], Daniel Engelhaupt[5], Iain Kerr[6], Roger Payne[6], C. Scott Baker[1,2,7]

[1] Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365, USA
[2] Department of Fisheries and Wildlife, Oregon State University, 104 Nash Hall, Corvallis, OR 97330, USA
[3] Biodiversity Institute, 1345 Jayhawk Blvd, University of Kansas, Lawrence, KS 66045, USA
[4] Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA
[5] HDR Inc., 1209 Independence Blvd, Virginia Beach, VA 23455, USA
[6] Ocean Alliance, 32 Horton Street, Gloucester MA 01930, USA

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.13638

This article is protected by copyright. All rights reserved

Author Manuscript

26 [7] School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142,
27 New Zealand
28
29 **Keywords:** Population genetics, population expansion, sex-biased dispersal, mtDNA,
30 microsatellite genotypes, Cetacea
31
32 * Corresponding author: Alana Alexander. Ph.: +785-864-9886; E-mail:
33 alana.alexander@.ku.edu
34
35 **Running title:** Genetic structure driven by female philopatry

36 **1. Abstract (287 words)**

37 The interplay of natural selection and genetic drift, influenced by geographic isolation, mating
38 systems, and population size, determines patterns of genetic diversity within species. The sperm
39 whale provides an interesting example of a long-lived species with few geographic barriers to
40 dispersal. Worldwide mtDNA diversity is relatively low, but highly structured among geographic
41 regions and social groups, attributed to female philopatry. However, it is unclear if this female
42 philopatry is due to geographic regions or social groups, or how this might vary on a worldwide
43 scale. To answer these questions, we combined mtDNA information for 1,091 previously
44 published samples with 542 newly obtained DNA profiles (394 bp mtDNA, sex, 13
45 microsatellites) including the previously un-sampled Indian Ocean, and social group information
46 for 541 individuals. We found low mtDNA diversity (π=0.430%) reflecting an expansion event
47 <80,000 years bp, but strong differentiation by ocean, among regions within some oceans, and
48 among social groups. In comparison, microsatellite differentiation was low at all levels,
49 presumably due to male-mediated gene flow. A hierarchical AMOVA showed that regions were
50 important for explaining mtDNA variance in the Indian Ocean, but not Pacific, with social group
51 sampling in the Atlantic too limited to include in analyses. Social groups were important in
52 partitioning mtDNA and microsatellite variance within both oceans. Therefore, both geographic
53 and social philopatry influence genetic structure in the sperm whale, but their relative importance
54 differs by sex and ocean, reflecting breeding behavior, geographic features, and perhaps a more
55 recent origin of sperm whales in the Pacific. By investigating the interplay of evolutionary forces

Author Manuscript

This article is protected by copyright. All rights reserved

operating at different temporal and geographic scales, we show that sperm whales are perhaps a unique example of a worldwide population expansion followed by rapid assortment due to female social organization.

2. Introduction

Despite the absence of obvious geographic barriers, striking patterns of genetic differentiation and diversity are evident in many marine megafauna. This includes low genetic diversity due to past population bottlenecks/expansions (e.g., giant squid, Winkelmann *et al.* 2013; killer whales, Moura *et al.* 2014); strong patterns of genetic differentiation due to prey specialization (e.g., killer whales, Riesch *et al.* 2012) or habitat specialization (e.g., harbor porpoises, Fontaine *et al.* 2014; sea lions, Lowther *et al.* 2012); genetic differentiation due to maternal or natal fidelity to breeding locations and migration routes (e.g., turtles, Bowen *et al.* 1992; baleen whales, Baker *et al.* 2013); and male-biased gene flow, as reflected in biparentally inherited nuclear markers and maternally inherited mitochondrial DNA (e.g., great white sharks, Pardini *et al.* 2001; humpback whales, Baker *et al.* 2013). In some species (e.g., killer whales, sperm whales), social groups also influence genetic differentiation, potentially reinforced by culture such as vocal dialects (Cantor *et al.* 2015; Cantor & Whitehead 2015; Gero *et al.* 2015; Hoelzel *et al.* 2007; Whitehead *et al.* 2012). While studies often investigate single factors that influence genetic diversity, teasing apart different mechanisms requires an assessment of genetic diversity patterns over multiple spatial and temporal time scales.

Due to its worldwide distribution (Gosho *et al.* 1984), social behavior (Gero *et al.* 2015), and acoustically mediated culture (Cantor *et al.* 2015; Cantor & Whitehead 2015), the sperm whale presents an interesting case study for this type of hierarchical analysis. Although whaling removed hundreds of thousands of individuals (Best 1979; Whitehead 2002; Whitehead 2003), the sperm whale is relatively abundant in comparison with other large whale species (~360,000 individuals worldwide; Whitehead 1998; 2002). Given the sperm whale's abundance and wide geographic range, mitochondrial DNA (mtDNA) diversity in sperm whales is relatively low compared with many other cetacean species (Alexander *et al.* 2013; Lyrholm *et al.* 1996; Whitehead 1998), yet marked by moderate-to-strong differentiation between oceans (Lyrholm & Gyllensten 1998), among marginal seas within the Atlantic (Drouot *et al.* 2004; Engelhaupt *et al.*

Author Manuscript

This article is protected by copyright. All rights reserved

86 2009), and among social groups within the Pacific (Lyrholm & Gyllensten 1998; Rendell *et al.*
87 2012). In an analysis of mitogenomes (Alexander *et al.* 2013), three previously proposed
88 hypotheses were considered as the most likely causes of the low mtDNA diversity in sperm
89 whales: a population bottleneck and/or expansion (Lyrholm & Gyllensten 1998; Lyrholm *et al.*
90 1996); a selective sweep due to a favorable substitution in a mtDNA-encoded protein (Janik
91 2001); or a selective sweep due to beneficial cultural traits transmitted matrilineally in parallel
92 with the mitogenome – cultural hitchhiking (Whitehead 1998; Whitehead 2005). In comparison,
93 although significant nuclear differentiation (based on microsatellite genotypes) has been observed
94 among social groups in the Pacific (Lyrholm *et al.* 1999), there is only weak differentiation
95 among regions within oceans (Engelhaupt *et al.* 2009; Mesnick *et al.* 2011), and no significant
96 nuclear differentiation between ocean basins (Lyrholm *et al.* 1999). The contrast between
97 mtDNA and microsatellite differentiation has been interpreted as male dispersal and female
98 philopatry at three hierarchical levels: oceanic scales (Lyrholm *et al.* 1999), between regions
99 within oceans (e.g., the Atlantic; Engelhaupt *et al.* 2009), and at the social group level (Lyrholm
100 & Gyllensten 1998).
101
102 Female philopatry and male-biased dispersal is consistent with behavioral observations of sperm
103 whale social structure. Males disperse from their natal social units at an age of 3-15 years (Best
104 1979; Richard *et al.* 1996a; Whitehead 2003) and become increasingly solitary as they age,
105 extending their latitudinal range into polar waters (Allen 1980; Best 1979; Whitehead 2003).
106 After reaching social maturity (at 25-27 years, Best 1979), males associate with females for the
107 purposes of mating, but do not permanently remain with any given female social group (Richard
108 *et al.* 1996a; Whitehead 1993; Whitehead 1994). Female social groups contain adult females that
109 show long-term social bonds with one another, as well as juveniles of both sexes, and are
110 confined to low-latitude tropical and temperate waters (Best 1979; Christal *et al.* 1998; Coakes &
111 Whitehead 2004; Dufault & Whitehead 1998; Dufault *et al.* 1999; Richard *et al.* 1996a).
112 However, there are substantial differences by ocean in vocal dialects, female social group size,
113 and proportion of calves within social groups (Gero *et al.* 2015; Whitehead *et al.* 2012),
114 suggesting that the relative importance of female social groups in partitioning genetic diversity
115 might vary by ocean.

Author Manuscript

This article is protected by copyright. All rights reserved

116

117 Here, we investigate the cause of the rapid radiation of maternal lineages in sperm whales, and
118 how social group and geographic factors partition genetic diversity in different oceans. We
119 hypothesize that the low but highly structured mtDNA diversity observed in the sperm whale is
120 consistent with a recent, rapid radiation of a single mtDNA lineage, followed by genetic drift due
121 to female philopatry at regional and social group levels. For this, we assembled the largest sperm
122 whale genetic dataset to date, including both published and previously unpublished data.
123 Previously unpublished data included stranding samples, and samples collected by the 'Voyage
124 of the *Odyssey*': a five year expedition that collected biopsy samples (skin and blubber) from
125 under-sampled equatorial regions (Figure 1), including the previously un-sampled Indian Ocean
126 (Godard *et al.* 2003). Previously published data included mtDNA control region (CR) haplotypes
127 from 1,091 samples (Engelhaupt *et al.* 2009; Lyrholm & Gyllensten 1998; Mesnick *et al.* 2011;
128 Rendell *et al.* 2012; Richard *et al.* 1996a; Whitehead *et al.* 1998). Using this unprecedented
129 dataset, we first explicitly test the hypothesis that a past population expansion could explain the
130 low mtDNA diversity seen in the sperm whale. We then evaluate the importance of geographic
131 regions versus social groups in determining genetic structure within the Pacific and Indian
132 Oceans (where sufficient individuals with social group information were available), using the
133 unique circumequatorial collection of samples from social groups and within-ocean regions.
134 Finally, we also examine sex-specific differences by carrying out these analyses for both
135 biparentally inherited nuclear markers (13 microsatellite loci) and the maternally inherited
136 mtDNA. This study demonstrates how different factors shape patterns of genetic diversity at
137 multiple scales in a broadly distributed marine organism.

138

139 **3. Materials and Methods**

140 ***3.1. Assembly of mtDNA dataset and definition of spatial scales***

141

142 Using the definitions developed by Mesnick *et al.* (2005, also see Supplementary Material 1), we
143 summarized mtDNA haplotype information from previous publications by ocean and within-
144 ocean region (Supplementary Material 2). Regions were defined by aggregating samples that
145 were obtained within ~500 km of each other, with the exception of the Mediterranean that was

Author Manuscript

This article is protected by copyright. All rights reserved

146 pooled over the entire sea for consistency with previous publications (Engelhaupt *et al.* 2009).
147 Areas included in regional analyses were restricted to those sampled by the *Odyssey* (and
148 augmented by samples from previous studies, where available), and tropical/subtropical areas
149 (38°S to 38°N) sampled in previous studies, as these were the latitudes primarily sampled by the
150 *Odyssey*. Regions were also required to have five or more sampled individuals to limit the effect
151 of low sample sizes. Aggregation of datasets from different publications was possible because of
152 the concerted efforts of the Cachalote Consortium (Mesnick *et al.* 2005) to standardize
153 nomenclature for sperm whale mtDNA CR haplotypes. A lack of standardized nuclear markers
154 did not allow for identification and removal of potential between-study replicates. However, we
155 removed within-study replicates where identified.
156
157 'Voyage of the *Odyssey*' samples were collected from 1999-2005 in circum-equatorial regions
158 (Figure 1) using a biopsy dart. As detailed in Godard *et al.* (2003), total genomic DNA was
159 extracted from the *Odyssey* samples using a high-salt procedure. DNA aliquots of 895 samples
160 were then provided by Ocean Alliance, sponsor of the 'Voyage of the *Odyssey*'. New Zealand
161 sperm whale skin and tissue samples (n = 89) were collected from strandings by New Zealand
162 Department of Conservation staff from 1994 to 2008 and archived in the New Zealand Cetacean
163 Tissue Archive (CeTA) at the University of Auckland. One sample originating in Samoa,
164 archived in CeTA, was also included. Oregon sperm whale skin and tissue samples (n = 3) were
165 provided by the Oregon Marine Mammal Stranding Network. DNA was extracted from samples
166 of stranded animals following a standard phenol/chloroform technique (Sambrook *et al.* 1989) as
167 modified by Baker *et al.* (1994).
168
169 We carried out amplification of the mtDNA CR using the primers M13dlp1.5 and tphe, and
170 sequenced a 619 bp consensus length of this fragment on an ABI3730xl DNA analyzer, as
171 described in Alexander *et al.* (2013). Sequences were trimmed using PHRED scores and by eye
172 in *Sequencher v. 4.6* (Gene Codes). After trimming, sequences with more than 10% of bases
173 showing a PHRED score of <20 were re-sequenced or removed from the dataset (Morin *et al.*
174 2010). We visually confirmed variable sites between haplotypes in each sequence using
175 *Sequencher*. After removal of replicates, we trimmed the *Odyssey* and stranding samples to the

Author Manuscript

This article is protected by copyright. All rights reserved

176 shorter consensus length of 394 bp and combined them with the previously published mtDNA
177 data. This 394 bp fragment has the highest level of diversity across the sperm whale mitogenome,
178 and accurately reflects intraspecific phylogenies based on the full mitogenome sequence
179 (Alexander *et al.* 2013).
180

181 ***3.2. Sex identification of* Odyssey *and stranding samples***
182

183 We sexed samples using a multiplexed PCR amplifying 152 bp of the *SRY* on the Y chromosome
184 of males (Richard *et al.* 1994; primers: sperm-whale specific *SRY* primers), and a 442-445 bp
185 fragment of the *ZFX/ZFY* fragment present in both males and females (Aasen and Medrano 1990;
186 primers: P1-5EZ and P2-3EZ). Each reaction consisted of 1 µL of sample DNA, and a final
187 concentration of 0.9× Platinum *Taq* buffer (Invitrogen), 0.36 µM of each of the four primers, 2.27
188 mM $MgCl_2$, 0.18 mM dNTP and 0.25 U of Platinum *Taq* polymerase (Invitrogen), with ddH_2O
189 to 11 µL total volume. The temperature profile consisted of an initial denaturing step of 3 min at
190 94°C, followed by 35 cycles of 94°C for 45 s, 58°C for 45 s and 72°C for 60 s, followed by a
191 final extension step of 72°C for 10 min. The PCR products were run on a 1.6% agarose gel
192 (buffer: TBE), stained with ethidium bromide, and visualized under UV light. Presence of two
193 bands indicated a male sample, one band a female sample, and no bands PCR failure.
194

195 ***3.3. Microsatellite genotyping, identification of replicates and kin***
196 Thirteen microsatellite loci were selected based on previous genotyping in the sperm whale
197 (Engelhaupt *et al.* 2009), and other cetacean species (Supplementary Material 3). We amplified
198 each locus in an individual reaction, with 1 µL of the sample DNA, a final concentration of 0.9×
199 Platinum *Taq* buffer (Invitrogen), 0.36 µM of each primer and 0.18 mM dNTP. $MgCl_2$ and
200 Platinum *Taq* polymerase (Invitrogen) concentrations varied by locus as detailed in
201 Supplementary Material 3, and we added ddH_2O to 11 µL total volume. Temperature profiles
202 consisted of an initial denaturing step of 3-5 minutes at 94-95°C, followed by 35-40 cycles of 94-
203 95°C for 30-40 s, the locus-specific annealing temperature (as detailed in Supplementary
204 Material 3) for 30-60 s, and 72°C for 30-60 s, followed by a final extension step of 72°C for 8-30
205 min. Multiple microsatellite loci were combined based on differing size range and fluorescent

This article is protected by copyright. All rights reserved

Author Manuscript

206 label (Supplementary Material 3) and co-loaded on an ABI3730xl DNA Analyzer with GS500 LIZ
207 ladder. Output was processed using *GeneMapper v. 3.7* (Applied Biosystems), with a minimum
208 signal strength detection threshold of 50 units. All automated calls were checked by eye, with a
209 subset of samples cross-checked by a second researcher (D. Steel) to ensure consistency in allele
210 calling.
211
212 Other quality control (QC) measures were then carried out following Morin *et al.* (2010).
213 Samples were only included if they were genotyped for at least eight microsatellite loci, and
214 successfully sexed or sequenced for mtDNA CR. This QC was to limit the inclusion of samples
215 likely affected by low DNA quantity/quality. Identification of replicate samples was carried out
216 using *Cervus v. 3.0* (Kalinowski *et al.* 2007). To compensate for genotyping error, we used
217 relaxed matching allowing for mismatches at up to four microsatellite loci, with mismatching loci
218 corrected or repeated. If remaining mismatches were consistent with allelic dropout, samples
219 were considered replicates if they matched at sex and mtDNA CR. Probability of identity ($p_{(ID)}$)
220 was calculated using *Cervus* for pairs showing exact matches, and *Genalex v. 6.501* (Peakall &
221 Smouse 2006; Peakall & Smouse 2012) for those with mismatches (average $p_{(ID)}$ for the
222 combination of exactly matching markers). The per-allele microsatellite error rate (Pompanon *et*
223 *al.* 2005) was estimated using intentional duplicates for 110 samples, selected randomly with
224 respect to DNA quality and quantity. Tests for deviation from Hardy Weinberg equilibrium and
225 for linkage disequilibrium were conducted using *Genepop v. 4.2* (Raymond & Rousset 1995a),
226 and tests for the presence of large allele dropout and null alleles using *MICRO-CHECKER v.*
227 *2.2.3* (van Oosterhout *et al.* 2004), following the methods of Carroll *et al.* (2011). We excluded
228 microsatellite loci from population genetic analyses if they showed departures from Hardy-
229 Weinberg equilibrium. A 'kin restricted' dataset was created to account for the sampling of first-
230 order relatives within social groups. We followed Mesnick *et al.*'s (2011) approach of removing
231 one member of every first-order kin pair (defined as samples sharing at least one allele at every
232 microsatellite locus) identified using *SOLOMON v. 1.0-1* (Christie *et al.* 2013). We retained the
233 sample with the most complete genotype from each pair. We conducted analyses on both the
234 'full' and 'restricted' version of this dataset.
235

Author Manuscript

This article is protected by copyright. All rights reserved

236 ***3.4. Testing for a population expansion***

237 We tested for a population expansion in the mtDNA dataset by assessing Fu's *F*s (Fu 1997) and
238 the mismatch distribution (Harpending 1994; Rogers & Harpending 1992; Slatkin & Hudson
239 1991) under a demographic expansion scenario through *Arlequin v. 3.5* with 10,000 permutations
240 to assess significance (Excoffier *et al.* 2005; Schneider & Excoffier 1999). We carried out these
241 tests at the worldwide, oceanic and regional levels.

242

243 ***3.5. mtDNA diversity, differentiation, and phylogeography***

244 Haplotype and nucleotide diversity (using the Tamura and Nei (1993) correction) were calculated
245 using *Arlequin*. Differences in haplotype diversity and nucleotide diversity between oceans were
246 assessed using a custom *R v. 3.0.2* (R Core Team 2013) script, to conduct a permutation test with
247 10,000 replicates (Alexander 2015). We inferred the number of substitutions between the mtDNA
248 CR haplotypes with a parsimony network created using *TCS v. 1.2.1* (Clement *et al.* 2000). We
249 examined oceanic differentiation using F_{ST} and Φ_{ST} (with the Tamura and Nei (1993) correction),
250 with 10,000 replicates to assess significance in *Arlequin*. To evaluate the potential influence of
251 phylogeographic structure (i.e., divergence as well as drift), we used *PERMUT v. 2.0* (Petit 2010)
252 to test for differences between G_{ST} and N_{ST} (analogs of F_{ST} and Φ_{ST}: Pons and Petit 1996).

253

254 ***3.6. Microsatellite diversity and differentiation***

255 Observed and expected heterozygosity were calculated using *Cervus*, and allelic richness using
256 *FSTAT v. 2.9.3* (Goudet 2001). We tested for significant differences in observed heterozygosity
257 and allelic richness between oceans using a custom *R* script (Hamner 2014) that implemented t-
258 tests or Wilcoxon signed rank tests depending on equality of variances and normality of
259 differences between areas (Supplementary Material 4). We calculated F_{IS} by region and ocean
260 (using oceanic and worldwide microsatellite allele frequencies, respectively) using *FSTAT*, with
261 10,000 replicates to assess significance. To examine oceanic differentiation, we calculated F_{ST}
262 through *Genepop* (Raymond & Rousset 1995b; Rousset 2008), using the exact test to assess
263 significance (Raymond & Rousset 1995a). G''_{ST}, an index that compensates for the diversity of
264 microsatellites (Meirmans & Hedrick 2011), was calculated with *Genodive v. 2.ob25* (Meirmans
265 & van Tienderen 2004), using 10,000 permutations to assess significance. The presence of

Author Manuscript

This article is protected by copyright. All rights reserved

266 population structure independent of our *a priori* partitions was assessed for the regional
267 microsatellite dataset using *Structure v 2.3.4* (Falush *et al.* 2003; Pritchard *et al.* 2000).
268 Following Engelhaupt *et al.* (2009), we assumed admixture and correlated allele frequencies with
269 500,000 burn-in steps, followed by 1,000,000 steps. Twenty replicates (following the
270 recommendations of Gilbert *et al.* 2012) were carried out for $K = 1$ to $K = 13$. The Evanno *et al.*
271 (2005) method was used to assess the best fitting K through *STRUCTURE HARVESTER* (Earl &
272 vonHoldt 2012). Using *STRUCTURE HARVESTER* output, *CLUMPP* (Jakobsson & Rosenberg
273 2007) was used to align cluster assignment across replicates.
274

275 ***3.7. Evaluating the relative importance of social group versus geographic regions by ocean***

276 We used field data on spatial and temporal proximity of *Odyssey* biopsy samples to identify
277 samples collected during a single encounter with a social group. To account for previously
278 sampled groups that were unintentionally re-encountered, we combined groups that had genetic
279 replicates between them. Our groups likely correspond to a mix of 'social groups' and 'social
280 units' as defined in previous publications (Christal *et al.* 1998; Gero *et al.* 2015; Whitehead
281 2003). We included social groups from the literature where mtDNA data were available
282 (Supplementary Material 5). Tests of genetic differentiation were conducted partitioning the
283 dataset by social group, for groups where two or more individuals passed QC. Hierarchical
284 analyses nesting social group within ocean and region at the worldwide scale, and within region
285 for the Pacific and Indian Oceans (where adequate numbers of groups were available) were
286 conducted through *Arlequin* for mtDNA (F_{ST} and Φ_{ST}), and *GDA v 1.0* (Lewis & Zaykin 2001)
287 for microsatellites (F_{ST} only). To limit the effect of small sample sizes, nested analyses were
288 restricted to a subset of social groups that had five or more individuals pass QC.
289

290 ***3.8. Testing for female philopatry: sex-biased gene flow and dispersal***

291 We restricted analyses of sex-biased gene flow and dispersal to samples genotyped in this study,
292 where sex information was available. Analyses were carried out at the oceanic and within-ocean
293 regional levels (including all oceans/regions with at least 2 individuals of each sex) following the
294 methods of Oremus *et al.* (2007). Due to limited numbers of social groups with two or more
295 sampled males (Supplementary Material 5), no analyses were conducted at the social group level.

This article is protected by copyright. All rights reserved

Author Manuscript

296 We investigated sex-biased dispersal using two methods in *FSTAT*: (1) a comparison of sex-
297 specific F_{ST} values for both mtDNA (coding the mtDNA CR as a homozygote locus) and
298 microsatellites and (2) calculation of the sex-specific variance of assignment index (vAIc) based
299 on microsatellites (Goudet *et al.* 2002). We tested the difference between sex-specific values
300 using 10,000 permutations. The more dispersive sex is expected to have a lower F_{ST} value
301 (method 1), but higher variance (method 2), than the more philopatric sex (Oremus *et al.* 2007).
302 We note that males in this dataset included immature males that had not dispersed from their
303 natal social group that could conservatively bias the tests against finding male-biased dispersal.
304 As well as sex-biased dispersal, we obtained sex-specific gene flow estimates using the formulas
305 presented in Hedrick *et al.* (2013) and microsatellite/mtDNA CR F_{ST} as the input values.
306

307 **4. Results**

308 ***4.1. Assembly of mtDNA dataset***

309 We summarized sequence information for 1,091 samples from previous studies (Engelhaupt *et al.*
310 2009; Lyrholm & Gyllensten 1998; Mesnick *et al.* 2011; Rendell *et al.* 2012; Richard *et al.*
311 1996a; Whitehead *et al.* 1998; as detailed in Supplementary Material 2). After removal of
312 replicates identified by genotyping, mtDNA sequences were available for 496 individuals in the
313 collection from the *Odyssey* and from strandings. These sequences were trimmed to a consensus
314 sequence length of 394 bp and combined with the previously published information resulting in
315 1,587 sequences included in analyses of mtDNA differentiation and diversity at the worldwide
316 and oceanic level. Of these, 998 samples were included in analyses of 16 regions within oceans
317 (Table 1).
318

319 ***4.2. Assembly of microsatellite dataset and quality control***

320 Of the 988 total samples genotyped in this study, 671 passed quality control, with a minimum of
321 8 microsatellite loci each. On average, the samples passing QC had microsatellite genotypes that
322 were 92.8% complete (S.D. = 9.58%), representing 12 out of a potentially complete genotype of
323 13 microsatellite loci. We identified replicates using between 6 and 13 overlapping loci with
324 $p_{(ID)}$s between 3.39E-21 and 1.76E-06, and $p_{(ID\text{-}sibs)}$ between 2.16E-06 and 4.50E-03. The per-

This article is protected by copyright. All rights reserved

Author Manuscript

325 allele microsatellite error rate was 1.27% based on 74 intentional duplicate pairs that passed QC.
326 This error was largely due to allelic drop out (>95%) that was then identified and corrected, and
327 was similar in magnitude to previous studies on sperm whales and other cetaceans (Baker *et al.*
328 2013; Carroll *et al.* 2011; Mesnick *et al.* 2011). Using the known duplicates, there was no
329 detectable error in designation of mtDNA haplotypes (i.e., an error rate of <0.7%), and only one
330 male/female discrepancy between a duplicate pair (e.g., an error rate of 1.69%).
331
332 After removal of replicates, the 671 genotypes that passed QC represented 542 individuals. Using
333 *SOLOMON* to identify pairs of individuals sharing an allele at every locus, we found 12 likely
334 first-order kin relationships. One pair from the Chagos Archipelago consisted of two males, with
335 the remaining relationships involving at least one female. Given the small number of identified
336 potential-first order kin, results for the 'full' and 'kin-restricted' datasets were very similar for all
337 analyses. Consequently, we provide results of the 'restricted' dataset only as Supplementary
338 Materials 6-7. We defined 13 within-ocean regions in the microsatellite dataset (Table 1). In
339 contrast with the mtDNA dataset, three regions were not represented as they were not genotyped
340 in this current study/had insufficient numbers of individuals pass QC: Hawai'i, western North
341 Atlantic, and the Gulf of Mexico. We did not detect consistent significant deviations from Hardy-
342 Weinberg equilibrium or linkage disequilibrium across the within-ocean regions. In addition, we
343 found no consistent evidence of scoring error, preferential large-allele dropout, or null alleles for
344 any microsatellite loci across the within-ocean regions. Therefore, we retained all loci (Table 2)
345 for analyses of microsatellite diversity and differentiation.
346
347 ***4.3. Testing for a population expansion***
348 For the worldwide dataset, Fu's Fs was strongly and significantly negative (Fs = -25.4, p =
349 0.0002), and the mtDNA mismatch distribution appeared unimodal (Figure 2; parameter
350 estimates and p-values for all comparisons listed in Supplementary Material 8). Along with a
351 star-like mtDNA network (Figure 3), these results are strong indicators of a worldwide
352 population expansion in female lineages of the sperm whale (Slatkin & Hudson 1991). Based on
353 τ = 1.625 and a control region substitution rate of 2.6%/million years (Alexander *et al.* 2013), the
354 age of the expansion event was estimated at 78,300 years ago (95% CI: 72,300 – 97,900),

This article is protected by copyright. All rights reserved

Author Manuscript

355 although the SSD and raggedness indices suggest the expansion model simulated by *Arlequin*
356 was not a good fit for the data ($p < 0.0109$). Because inference of population expansions can be
357 affected by population structure (Pannell & Whitlock 2003; Ptak & Przeworski 2002), we also
358 looked for population expansions at the oceanic and within-ocean regional levels. The Pacific
359 Ocean had a strongly significant negative Fu's Fs value (Fs = -21.5, $p = 0.0003$), and both the
360 Pacific and Atlantic gave qualitatively unimodal mismatch distributions (Figure 2), with the
361 population expansion model supported for the Atlantic Ocean ($p > 0.1547$). In contrast, the
362 Indian Ocean showed a multi-modal mismatch distribution (Figure 2) and the Fu's Fs value was
363 not significant (Supplementary Material 8). Using estimates of tau for each ocean, the time at
364 expansion within the Pacific was estimated at 66,900 years before present (95% CI: 60,800 –
365 87,300); 67,200 years before present in the Atlantic (95% CI: 55,200 – 86,700), and 94,000 years
366 before present in the Indian Ocean (95% CI: 37,100 – 150,000). These data suggest a more recent
367 population expansion event in the Pacific, also supported by the large number of within-Pacific
368 regions with negative Fu's Fs results (Supplementary Material 8).
369
370 ***4.4. mtDNA diversity, differentiation, and phylogeography***
371 We resolved a total of 39 mtDNA CR haplotypes in the worldwide dataset (Table 1), including
372 twelve previously unreported haplotypes (Figure 3). Except for *KK*, these new haplotypes were
373 rare ($n < 5$) and only found in one region (Figure 4). The maximum distance between any two
374 haplotypes was two substitutions and this only occurred twice on the haplotype network (Figure
375 3). Of the 31 variable sites found over the 394 bp mtDNA CR, all were transitions
376 (Supplementary Material 1). To investigate the potential for resolving further mtDNA diversity,
377 we sequenced 400 samples for 619 bp of the mtDNA CR. A comparison of the two consensus
378 lengths indicated 394 bp captured the majority of variation (Supplementary Material 6).
379 Therefore, even with the addition of Indian Ocean samples, the level of mtDNA CR diversity in
380 the sperm whale is still among the lowest in Cetacea (Table 1 vs. cetacean mtDNA diversity
381 estimates in Alexander *et al.* 2013). The Atlantic Ocean had significantly lower nucleotide
382 diversity than the Indian and Pacific Oceans, and significantly lower haplotype diversity than the
383 Pacific (Table 1, *p*-values for all significant diversity comparisons summarized in Supplementary

Author Manuscript

This article is protected by copyright. All rights reserved

384 Material 9), which appeared to be partly driven by the lack of mtDNA CR variation within the
385 Mediterranean (Table 1).
386
387 There was some sharing of mtDNA haplotypes across all three ocean basins, particularly of *A*, *B*,
388 and *C*: the three most common haplotypes (Figure 3, Figure 4). Despite this, geographic structure
389 was evident at the oceanic scale, with four haplotypes found at reasonably high frequencies (in
390 >20 individuals) yet restricted to a single ocean basin (haplotype *X* in the Atlantic, *KK* in the
391 Indian Ocean, *E* and *D* in the Pacific Ocean: Figure 3, Figure 4). Values for F_{ST} and Φ_{ST} showed
392 similar patterns at oceanic and regional levels (Figure 4, Supplementary Material 10), albeit with
393 Φ_{ST} tending to exceed the magnitude of F_{ST}. An explicit test of these two indices (using N_{ST} as
394 an analog of Φ_{ST}; Pons and Petit, 1996), indicated a small but significant ($p < 0.05$) influence of
395 phylogeographic structure at the oceanic level (i.e., haplotype lineages sorted by ocean) and
396 worldwide regional level (i.e., haplotype lineages sorted by region over a worldwide scale).
397 However, these results were contingent on the inclusion of the Gulf of Mexico, which has the
398 closely related haplotypes X and Y present in high frequencies (Figure 3, Figure 4).
399
400 ***4.5. Microsatellite diversity and differentiation***
401 No significant differences in microsatellite heterozygosity or allelic richness were found between
402 oceans (using Wilcoxon signed rank or t-tests depending on equality of variances and normality
403 of differences between pairs; locus by locus results Supplementary Material 7). Significant
404 differentiation was detected among oceans (F_{ST} 0.003, G''_{ST} 0.015, $p < 0.05$), but this was far
405 lower in magnitude than that seen for mtDNA (Figure 4). Our *Structure* results gave the highest
406 likelihood to $K = 1$ (mean Ln Pr(X|*K*) = -18,855.3 *cf.* -18,908.9 for $K = 2$). Visual inspection of
407 the structure results for $K = 2$ (Supplementary Material 11) showed no obvious population
408 structure, offering further support for $K = 1$. This is not surprising given the low levels of
409 differentiation found in the *a priori* analyses partitioning the dataset by region (e.g., $F_{ST} < 0.02$:
410 Supplementary Material 10, 12) (Waples & Gaggiotti 2006).
411
412 ***4.6. Evaluating the relative importance of social group versus geographic regions by ocean***

Author Manuscript

This article is protected by copyright. All rights reserved

413 Among the *Odyssey* dataset, 67 social groups (*n* = 420 individuals) had more than two
414 individuals pass genetic QC measures. After inclusion of published mtDNA information for 28
415 social groups from Ecuador and the Gulf of Mexico (Engelhaupt 2004; Ortega-Ortiz *et al.* 2012;
416 Richard *et al.* 1996a), we included 95 social groups (*n* = 541 individuals), representing 16
417 regions and all 3 oceans (Supplementary Material 5). In non-nested analyses, differentiation
418 among social groups was extremely high: social group consistently explained greater levels of
419 variation than partitioning by regions and oceans (Figure 4). This is expected, given the more
420 fine-scale partitioning of social groups compared to higher-level geographic scales.
421
422 To account for this fine-scale partitioning in a hierarchical AMOVA, we nested social groups
423 with ≥ 5 sampled individuals within ocean and region for the worldwide dataset, and within
424 regions for the Pacific and Indian Oceans (the Atlantic did not have enough regions for each
425 genetic marker type with social groups of five or more individuals). For the worldwide dataset,
426 social group explained a greater amount of mtDNA variance than either ocean or region, but all
427 levels were significant (Table 3). In the Pacific, only social group (compared with region)
428 explained any significant amount of mtDNA variance (Table 3). In the Indian Ocean, however,
429 region explained a larger percentage of variance than social group, although social group was
430 significant (Table 3). Social group was the only hierarchical level that explained any significant
431 variance in the microsatellite dataset (Table 3).
432

433 ***4.7. Testing for female philopatry: sex-biased gene flow and dispersal***

434 Most equatorial regions showed a significant skew towards females, consistent with the
435 assumption that the *Odyssey* largely targeted social groups dominated by females (Table 1). Sex-
436 specific estimates of gene flow, calculated from microsatellite and mtDNA (Hedrick *et al.* 2013),
437 were low for females and high for males (Table 4). Given the evidence for sex-biased gene flow,
438 it is not surprising that tests for sex-biased dispersal indicated males are the more dispersive sex
439 (Table 4). At all hierarchical levels, female-specific F_{ST} for mtDNA exceeded that of males, and
440 was significantly greater when partitioning by regions over the worldwide dataset (Table 4). For
441 microsatellites, the magnitude of female-specific F_{ST} was greater than male-specific
442 differentiation except among regions within the Pacific (however, neither sex's F_{ST} was

This article is protected by copyright. All rights reserved

Author Manuscript

443 significantly different from zero). Female-specific microsatellite F_{ST} significantly exceeded that
444 of males at the oceanic scale (Table 4). Surprisingly, the sex-specific variance of assignment tests
445 were not significant.
446
447 **5. Discussion**
448 We have shown evidence for multiple forces operating on genetic diversity and differentiation in
449 the sperm whale, a marine species with a worldwide distribution, over different temporal and
450 geographic scales. We suggest the relatively low mtDNA diversity of sperm whales is consistent
451 with a recent population expansion or sweep. However, despite the low mtDNA diversity, we
452 detected marked patterns of maternal structure in the Indian Ocean, similar to that seen in the
453 Atlantic Ocean (this study; Engelhaupt *et al.* 2009), but in the absence of obvious geographic
454 boundaries. In contrast, the Pacific Ocean showed far less regional mtDNA differentiation. Even
455 after accounting for social group in a nested AMOVA, region remained an important level in
456 describing genetic structure within the Indian Ocean, but not the Pacific. This is consistent with
457 previous studies that found no geographically based mtDNA structure in the Pacific (Lyrholm &
458 Gyllensten 1998; Lyrholm *et al.* 1999; Rendell *et al.* 2012; Whitehead *et al.* 1998), or significant,
459 but low levels of differentiation (Mesnick *et al.* 2011), suggesting the Pacific is unusual in its
460 lack of geographic structure in comparison with the Atlantic and Indian Oceans. In contrast with
461 the high levels of maternal structure found at various hierarchical scales, nuclear structure was far
462 less pronounced (albeit significant at the oceanic level, in contrast with previous studies e.g.,
463 Lyrholm *et al.* 1999, likely due to our larger sample sizes). In fact, within the nested AMOVA,
464 social group was the only important level for describing microsatellite variance. Although the
465 lack of nuclear structure could be influenced by the recent population expansion/sweep, it is also
466 likely affected by the presence of male-biased dispersal and gene flow.
467
468 ***5.1. A recent worldwide expansion of sperm whales***
469 The mismatch analysis conducted in this study was consistent with a worldwide expansion of a
470 single maternal lineage that began ~80,000 years ago. It is important to point out that this
471 estimate is provisional due to the problems of model fitting and phylogenetically derived
472 substitution rates (Grant 2015; Ho *et al.* 2011a). Indeed, the use of a faster substitution rate

This article is protected by copyright. All rights reserved
Author Manuscript

473 derived from ancient DNA sampling (e.g., 20%/million year for bowhead whales: Ho *et al.*
474 2011b) would lead to an estimate of the expansion beginning ~10,000 years ago, consistent with
475 the end of the last glacial maximum (LGM) (Lambeck *et al.* 2014). In a remarkable parallel,
476 another abyssal predator (and one of the sperm whale's prey), the giant squid (*Architeuthis* spp.),
477 also shows extremely low mitogenomic diversity (Winkelmann *et al.* 2013), and a similar time to
478 most recent common maternal ancestor, depending on the substitution rate used. This raises the
479 possibility that a worldwide expansion of sperm whales could have been predicated on a recent
480 expansion of their prey, especially as other squid species have also shown signatures of
481 demographic/range expansions that appear to be associated with the LGM (e.g., *Doryteuthis gahi*,
482 Ibáñez *et al.* 2012; *Dosidicus gigas,* Ibáñez *et al.* 2011; Ibáñez & Poulin 2014). Further support
483 for this hypothesis comes from other deep-diving, squid-feeding cetaceans which show similar
484 patterns of expansion, including the gray's beaked whale (*Mesoplodon grayi*) and pilot whales
485 (*Globicephala* spp.) (Oremus *et al.* 2009; Thompson *et al.* 2016). However, as the squid species
486 mentioned are only some of the many cephalopod and fish species preyed on by the sperm whale
487 (Whitehead 2003), future research should establish whether other prey species show the same
488 patterns. This is especially pertinent as other cetacean species with diverse prey bases have also
489 shown signatures of population expansions associated with the LGM (e.g., white-beaked
490 dolphins, Banguera-Hinestroza *et al.* 2010; harbor porpoises, Fontaine *et al.* 2014; killer whales,
491 Morin *et al.* 2015).
492
493 In addition, the population expansions of the squid species mentioned could also support the
494 cultural hitchhiking hypothesis, if the use of these squid as prey was restricted to a few initial
495 sperm whale matrilines. The inclusion of population-level nuclear genetic markers in future
496 studies could distinguish between a selective sweep and a population expansion as the cause of
497 low mtDNA diversity. A selective sweep (either due to cultural hitchhiking, or functional
498 selection acting on the mtDNA) will reduce the genetic diversity of the mitogenome, but not of
499 the nuclear genome (Charlesworth *et al.* 2003; Rokas *et al.* 2001). Given the limited
500 phylogeographic structure (i.e., divergence) observed for sperm whale mtDNA, whatever the
501 ultimate cause of the low mtDNA diversity, time since this point has been insufficient for unique
502 ocean-specific or region-specific haplotype lineages (with the exception of the Gulf of Mexico) to

This article is protected by copyright. All rights reserved

Author Manuscript

503 be established through mutation and lineage sorting. However, the marked female philopatry
504 present in the sperm whale at regional and social group levels has worked on post-expansion
505 mtDNA diversity to establish strong patterns of mtDNA differentiation within oceans.
506
507 ***5.2. Female philopatry at the geographic vs social group level varies by ocean***
508
509 Levels of regional differentiation in mtDNA were much higher in the Atlantic and Indian Oceans
510 than in the Pacific. Geographic region persisted as an important factor in partitioning mtDNA
511 diversity within the Indian Ocean, even after accounting for social group. In contrast, within the
512 Pacific, social group was the only level that described any significant amount of variation. The
513 lack of regional structure in the Pacific is consistent with behavioral evidence: female whales in
514 the Pacific appear to range further than in the Atlantic, up to ~4000 km in the Pacific and only up
515 to ~700 km in the Atlantic (Jaquet *et al.* 2003; Mizroch & Rice 2013; Ortega-Ortiz *et al.* 2012;
516 Whitehead *et al.* 2012; Whitehead *et al.* 2008). Previously, differences in geographic structure
517 and social group composition between the Atlantic and Pacific Oceans have been attributed to
518 oceanography, predation, whaling or culture (Whitehead et al. 2012). Our results suggest that a
519 consideration of the factors driving differences in geographic structure should also be extended to
520 the Indian Ocean.
521
522 Oceanography can influence differentiation through geographic isolation (e.g., the
523 Mediterranean, Gulf of Mexico: Engelhaupt *et al.* 2009). However, geographic isolation cannot
524 explain the large degree of mtDNA differentiation seen within the Indian Ocean, particularly
525 illustrated by the mtDNA haplotype frequency differences in comparison with Sri Lanka. Instead
526 a potential oceanographic explanation lies in the bathymetry of Sri Lanka: there are a large
527 number of submarine canyons that lead to enhanced productivity of this region (de Vos *et al.*
528 2012). Female sperm whales utilizing the Sri Lankan canyons might not need to range as widely
529 to satisfy nutritional requirements (Gordon 1987; Moors-Murphy 2014). This 'enhanced
530 philopatry' could then lead to the striking geographic differentiation in mtDNA observed. A
531 similar process of local fidelity has been proposed for insular communities of otherwise pelagic
532 dolphins, due to an 'island mass' effect (Martien *et al.* 2012; Oremus *et al.* 2012).

This article is protected by copyright. All rights reserved

Author Manuscript

533

534 Another potential explanation for the difference in geographic structure between oceans lies in
535 the acoustic culture of sperm whales. It has been previously hypothesized that acoustic clans,
536 which comprise of social groups with similar repertoires of acoustic codas (stereotypical series of
537 clicks), shape patterns of genetic differentiation in the sperm whale (Rendell *et al.* 2012; Watkins
538 & Schevill 1977; Whitehead *et al.* 1998). In the Atlantic, coda patterns vary based on geographic
539 regions and acoustic clans are allopatric (Whitehead *et al.* 2012). This correlates with the
540 heightened patterns of geographically based mtDNA differentiation seen in this ocean. In the
541 Pacific, acoustic clans are distributed sympatrically across broad geographic ranges (Cantor *et al.*
542 2015; Rendell *et al.* 2012; Whitehead *et al.* 1998). It has been previously proposed that the lack
543 of geographically based mtDNA differentiation in the Pacific is because maternal dispersal and
544 gene flow occurs within acoustic clans, but across broad geographic scales (i.e., females are
545 socially philopatric rather than geographically philopatric; Cantor *et al.* 2015; Rendell *et al.*
546 2012; Whitehead *et al.* 1998). This hypothesis is consistent with our nested AMOVA results for
547 the Pacific, where social group was the only level that explained any significant amount of
548 genetic variation. However, genetic structure driven by oceanography or culture are not
549 necessarily mutually exclusive hypotheses: differences in resource use could be reinforced by
550 differences in coda repertoire between acoustic clans (Cantor *et al.* 2015; Cantor & Whitehead
551 2015; Gero *et al.* 2015).

552

553 Local population declines due to whaling could have also reduced geographic structure in the
554 Pacific (Whitehead *et al.* 2012). Large-scale movements from the Galapagos to Peru have been
555 documented, where whaling-related declines left the productive Humboldt Current
556 underpopulated (Whitehead *et al.* 1997). Indeed, the Pacific, particularly the North Pacific, was
557 subjected to high levels of both legal and illegal whaling (Ivashchenko *et al.* 2013). Alternately,
558 perhaps there has been insufficient time in the Pacific for geographic structure to evolve, for
559 either genetic diversity or vocalization patterns, given our results suggested a more recent
560 expansion in the Pacific. Distinguishing between whaling and pre-human causes of the Pacific-
561 wide expansion will require additional genetic data to establish tighter confidence-intervals on

Author Manuscript

This article is protected by copyright. All rights reserved

562 the relative timing of population expansions between the oceans. Whether the signature of a
563 population expansion has been exacerbated by whaling or not, an expansion would be expected to
564 tightly correlate both the maternally inherited mtDNA and maternally influenced coda type
565 (Cantor *et al.* 2015; Whitehead *et al.* 2012). This could be further tested using linked acoustic and
566 genetic sampling in the Indian Ocean (e.g., Rendell *et al.* 2012). We would predict that in the
567 Indian Ocean, mtDNA genetic variation would be at equilibrium with both coda and geographic
568 structure, as it is in the Atlantic.
569

570 ***5.3. Female philopatry and male-biased dispersal***

571 Our findings confirm the importance of female philopatry and male-biased dispersal in the sperm
572 whale (Engelhaupt *et al.* 2009; Lyrholm *et al.* 1999). In addition to sex-biased dispersal, we
573 demonstrated that the sperm whale shows male-biased gene flow. Male-biased gene flow could
574 explain the significant among-group microsatellite differentiation detected in this study: differing
575 paternal contributions to the alleles present within each female social group would enhance
576 nuclear genetic drift between social groups (Richard *et al.* 1996a). This mechanism could be
577 investigated in the future using a gametic mark-recapture framework to detect paternities among
578 different social groups (Carroll *et al.* 2012; Garrigue *et al.* 2004). However, despite the overall
579 patterns of male-biased dispersal and gene flow, significant microsatellite differentiation between
580 oceans indicates some restriction in oceanic dispersal and gene flow, even of males. There is also
581 some evidence for breeding fidelity of males at even finer spatial scales than at the oceanic level,
582 as suggested by a possible first-order kinship between two males in the Chagos Archipelago.
583 Evidence for male fidelity has also been found in the Californian Current by Mesnick *et al.*
584 (2011) using genetic assignment. These findings could indicate that sex-biased dispersal in sperm
585 whales is facultative rather than obligate e.g., some males show philopatry to specific areas while
586 others disperse. A re-examination of other species (e.g., great white sharks, Pardini *et al.* 2001;
587 humpback whales, Baker *et al.* 2013) that show apparent signatures of male-biased gene flow
588 could be of interest to establish whether this phenomenon is found in other taxa.
589

590 ***5.4. Management implications***

Author Manuscript

This article is protected by copyright. All rights reserved

591 Sex-biased dispersal and strong maternal population structure in the sperm whale argues for
592 management units based on the more philopatric females, rather than the wider ranging males,
593 requiring female-specific estimates of population size similar to male-specific effective
594 population size estimates in humpbacks (Constantine *et al.* 2012). In addition, when defining
595 female-based population structure in the sperm whale, it is important to aggregate samples at
596 appropriate spatial scales (Donovan 1991; Dufault *et al.* 1999). Given the clustered sample
597 collection of the *Odyssey*, we chose to group samples that occurred within 500 km of another
598 sample. This could have inadvertently either split regions that were truly one population, or
599 alternately 'lumped' areas with more than one distinct population. Both of these alternatives
600 present problems. 'Splitting' regional populations could mean that the strong differences between
601 social groups detected in our current research, as well as previous studies (Lyrholm & Gyllensten
602 1998; Rendell *et al.* 2012), are conflated with regional differentiation. The alternative of
603 'clumping' can also be problematic as regions could represent areas of different importance for
604 males and females (e.g., the Maldives/Chagos Archipelago region, where only males were
605 sampled around the Chagos Archipelago, but both sexes around the Maldives). Although we
606 attempted to assess these *a priori* divisions against a non-*a priori* clustering method (*Structure*),
607 this was limited by low levels of differentiation and the relatively small number of microsatellite
608 loci.
609
610 Continuing to investigate patterns of genetic differentiation in the sperm whale is important, as
611 high degrees of differentiation (i.e., isolation) could indicate susceptibility to population declines
612 resulting from various ongoing anthropogenic threats, including: entanglement in marine debris
613 and ship strike (Notarbartolo-Di-Sciara 2014); exposure to pollutants (Savery *et al.* 2013a;
614 Savery *et al.* 2013b; Wise Sr *et al.* 2009; Wise Sr. *et al.* 2011), as well as pollution associated
615 with oil extraction (e.g., 2010 *Deepwater Horizon* oil spill), and negative interactions with
616 anthropogenic sound (Mate *et al.* 1994).
617
618 ***5.5. Conclusion***
619 In this study, we demonstrated that low mtDNA diversity in the sperm whale is likely due to a
620 recent population expansion. Despite low mtDNA diversity, we demonstrated high levels of

This article is protected by copyright. All rights reserved

Author Manuscript

621 regional structure within some ocean basins. However, social group was also an important level
622 in describing mtDNA variance. The importance of social group and geographic philopatry
623 differed by ocean, with only social group explaining any significant amount of mtDNA variance
624 in the Pacific Ocean. Being able to distinguish between geographic and social group philopatry is
625 important as a restriction in movement between local populations could indicate that there is a
626 real risk of long-term declines in response to current anthropogenic threats, despite the sperm
627 whale's large worldwide population size. The approach we have used in this study for
628 partitioning the effects of social group and geographic regions will also be useful for other
629 species that show strong social structure, yet are of conservation concern, such as elephants, the
630 long-finned pilot whale, and the killer whale (Archie *et al.* 2008; Hoelzel *et al.* 2007;
631 Ottensmeyer & Whitehead 2003).
632
633 The specific mechanism(s) driving the differing contributions to genetic structure within oceans
634 require further study: particularly whether the expansion within the Pacific is more recent than in
635 other oceans, and how acoustic codas are structured in the Indian Ocean. However, overall, the
636 high levels of mtDNA structure observed in the sperm whale appear to be driven by female
637 philopatry at multiple hierarchical levels, contrasting with male-biased dispersal and gene flow.
638 By investigating the interplay of evolutionary forces operating at different temporal and
639 geographic scales, we have shown that sperm whales are perhaps a unique example of a global
640 population expansion followed by rapid assortment due to female social organization.
641

642 **6. Acknowledgements**

643 Genetic samples from the 'Voyage of the *Odyssey*' were collected under permit #0751-1614 from
644 the US National Marine Fisheries Service. Principle Investigator: Iain Kerr. The authors thank all
645 who served as staff and crew during the Voyage of the *Odyssey*, as well as J. Wise Sr. and C.
646 LaCerte for curation of DNA from samples collected by the *Odyssey*. We thank R. Constantine
647 and K. Thompson for curation of the CeTA database, and New Zealand Department of
648 Conservation staff for collecting NZ samples used in this study; J. Rice for curation of the
649 OMMSN database that supplied the Oregon samples; Oregon State University Cetacean
650 Conservation and Genomics Laboratory for additional lab support; C. Sislak for assistance with

Author Manuscript

This article is protected by copyright. All rights reserved

651 DNA extraction, S. Pierszalowski and R. Hamner for analysis recommendations and assistance,
652 and M. Smith for assistance summarizing *Odyssey* toxicology results. We thank E. Carroll, A.
653 Liston, K. O'Malley, B. Taylor, R. Glor; S. Palumbi; L. Rieseberg and three anonymous
654 reviewers for valuable comments on this manuscript. This work was supported by a Mamie
655 Markham Award and a Lylian Brucefield Reynolds Award from the Hatfield Marine Science
656 Center; a 2008–2011 International Fulbright Science & Technology award to A.A.; and co-
657 funded by the ASSURE program of the Department of Defense in partnership with the National
658 Science Foundation REU Site program to K.H. and C.S.B. [grant number NSF OCE-1004947].
659 Publication of this paper was supported, in part, by the Thomas G. Scott Publication Fund.

660 **7. References**

661

662 Aasen E, Medrano JF (1990) Amplification of the ZFY and ZFX genes for sex identification in
663 humans, cattle, sheep and goats. *Nature Biotechnology* **8**, 1279-1281.
664 Alexander A (2015) genetic_diversity_diffs v1.0.0. Available from
665 https://github.com/laninsky/genetic_diversity_diffs.
666 Alexander A, Steel D, Slikas B*, et al.* (2013) Low diversity in the mitogenome of sperm whales
667 revealed by next-generation sequencing. *Genome Biology and Evolution* **5**, 113–129.
668 Allen KR (1980) Size distribution of male sperm whales in the pelagic catches. *Report of the*
669 *International Whaling Commission, Special Issue 2.* **SC/SPC/9**, 51-56.
670 Archie EA, Maldonado JE, Hollister-Smith JA*, et al.* (2008) Fine-scale population genetic
671 structure in a fission–fusion society. *Molecular Ecology* **17**, 2666-2679.
672 Baker CS, Slade RW, Bannister JL*, et al.* (1994) Hierarchical structure of mitochondrial DNA
673 gene flow among humpback whales *Megaptera novaeangliae,* world-wide. *Molecular*
674 *Ecology* **3**, 313-327.
675 Baker CS, Steel D, Calambokidis J*, et al.* (2013) Strong maternal fidelity and natal philopatry
676 shape genetic structure in North Pacific humpback whales. *Marine Ecology Progress*
677 *Series* **494**, 291-306.
678 Banguera-Hinestroza E, Bjørge A, Reid R, Jepson P, Hoelzel A (2010) The influence of glacial
679 epochs and habitat dependence on the diversity and phylogeography of a coastal dolphin
680 species: *Lagenorhynchus albirostris*. *Conserv Genet* **11**, 1823-1836.

This article is protected by copyright. All rights reserved

Author Manuscript

681 Bérubé M, Jørgensen H, McEwing R, Palsbøll PJ (2000) Polymorphic di-nucleotide
682 microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*.
683 *Molecular Ecology* **9**, 2181-2183.
684 Best PB (1979) Chapter 7: Social organization in sperm whales, *Physeter macrocephalus*. In:
685 *Behavior of Marine Animals. Volume 3: Cetaceans* (eds. Winn HE, Olla BL), pp. 277-
686 289. Plenum Press, New York-London.
687 Bowen BW, Meylan AB, Ross JP*, et al.* (1992) Global population structure and natural history of
688 the green turtle (*Chelonia mydas*) in terms of matriarchal phylogeny. *Evolution* **46**, 865-
689 881.
690 Buchanan FC, Friesen MK, Littlejohn RP, Clayton JW (1996) Microsatellites from the beluga
691 whale *Delphinapterus leucas*. *Molecular Ecology* **5**, 571-575.
692 Cantor M, Shoemaker LG, Cabral RB*, et al.* (2015) Multilevel animal societies can emerge from
693 cultural transmission. *Nat Commun* **6**.
694 Cantor M, Whitehead H (2015) How does social behavior differ among sperm whale clans?
695 *Marine Mammal Science*, n/a-n/a.
696 Carroll E, Patenaude NJ, Alexander A*, et al.* (2011) Population structure and individual
697 movement of southern right whales around New Zealand and Australia. *Marine Ecology*
698 *Progress Series* **432**, 257-268.
699 Carroll EL, Childerhouse SJ, Christie M*, et al.* (2012) Paternity assignment and demographic
700 closure in the New Zealand southern right whale. *Molecular Ecology* **21**, 3960-3973.
701 Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic
702 structure on neutral variation. *Annu Rev Ecol Evol S* **34**, 99-125.
703 Christal J, Whitehead H, Letteval E (1998) Sperm whale social units: variation and change.
704 *Canadian Journal of Zoology* **76**, 1431-1440.
705 Christie MR, Tennessen JA, Blouin MS (2013) Bayesian parentage analysis with systematic
706 accountability of genotyping error, missing data and false matching. *Bioinformatics* **29**,
707 725-732.
708 Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene
709 genealogies. *Molecular Ecology* **9**, 1657-1660.

Author Manuscript

This article is protected by copyright. All rights reserved

Author Manuscript

710 Coakes AK, Whitehead H (2004) Social structure and mating system of sperm whales off
711 northern Chile. *Canadian Journal of Zoology* **82**, 1360-1369.
712 Constantine R, Jackson JA, Steel D*, et al.* (2012) Abundance of humpback whales in Oceania
713 using photo-identification and microsatellite genotyping. *Marine Ecology Progress Series*
714 **453**, 249-261.
715 de Vos A, Clark R, Johnson C*, et al.* (2012) Cetacean sightings and acoustic detections in the
716 offshore waters of Sri Lanka: March-June 2003. *Journal of Cetacean Research and*
717 *Management* **12**, 185-193.
718 Donovan GP (1991) A review of IWC stock boundaries. *Report of the International Whaling*
719 *Commission (Special Issue 13)*, 39-68.
720 Drouot V, Bérubé M, Gannier A*, et al.* (2004) A note on genetic isolation of Mediterranean
721 sperm whales (*Physeter macrocephalus*) suggested by mitochondrial DNA. *Journal of*
722 *Cetacean Research and Management* **6**, 29-32.
723 Dufault S, Whitehead H (1998) Regional and group-level differences in fluke markings and
724 notches of sperm whales. *Journal of Mammalogy* **79**, 514-520.
725 Dufault S, Whitehead H, Dillon M (1999) An examination of the current knowledge on the stock
726 structure of sperm whales (*Physeter macrocephalus*) worldwide. *Journal of Cetacean*
727 *Research and Management* **1**, 1-10.
728 Duncan SI, Crespi EJ, Mattheus NM, Rissler LJ (2015) History matters more when explaining
729 genetic diversity within the context of the core–periphery hypothesis. *Molecular Ecology*
730 **24**, 4323-4336.
731 Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for
732 visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
733 *Genetics Resources* **4**, 359-361.
734 Engelhaupt D, Hoelzel AR, Nicholson C*, et al.* (2009) Female philopatry in coastal basins and
735 male dispersion across the North Atlantic in a highly mobile marine species, the sperm
736 whale (*Physeter macrocephalus*). *Molecular Ecology* **18**, 4193-4205.
737 Engelhaupt DT (2004) *Phylogeography, kinship and molecular ecology of sperm whales*
738 *(Physeter macrocephalus)* PhD, University of Durham.

This article is protected by copyright. All rights reserved

739 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the
740 software structure: a simulation study. *Molecular Ecology* **14**, 2611-2620.
741 Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: an integrated software package for
742 population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.
743 Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus
744 genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
745 Fontaine MC, Roland K, Calves I*, et al.* (2014) Postglacial climate changes and rise of three
746 ecotypes of harbour porpoises, *Phocoena phocoena*, in western Palearctic waters.
747 *Molecular Ecology* **23**, 3306-3321.
748 Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking
749 and background selection. *Genetics* **147**, 915-925.
750 Garrigue C, Dodemont R, Steel D, Baker CS (2004) Organismal and 'gametic' capture-recapture
751 using microsatellite genotyping confirm low abundance and reproductive autonomy of
752 humpback whales on the wintering grounds of New Caledonia. *Marine Ecology Progress*
753 *Series* **274**, 251-262.
754 Gero S, Gordon J, Whitehead H (2015) Individualized social preferences and long-term social
755 fidelity between social units of sperm whales. *Animal Behaviour* **102**, 15-23.
756 Gilbert KJ, Andrew RL, Bock DG*, et al.* (2012) Recommendations for utilizing and reporting
757 population genetic analyses: the reproducibility of genetic clustering using the program
758 structure. *Molecular Ecology* **21**, 4925-4930.
759 Godard C, Clark R, Kerr I, Madsen PT, Payne R (2003) Preliminary report on the sperm whale
760 data collected during the *Voyage of the Odyssey*. *Paper SC/55/O17 presented to the IWC*
761 *Scientific Committee, May 2003, Berlin*, 8pp.
762 Gordon JCD (1987) Sperm whale groups and social behavior observed off Sri Lanka. *Report of*
763 *the International Whaling Commission* **37**, 205-217.
764 Gosho ME, Rice DW, Breiwick JM (1984) The sperm whale, *Physeter macrocephalus*. *Marine*
765 *Fisheries Review* **46**, 54-64.
766 Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices
767 (version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm.

This article is protected by copyright. All rights reserved

Author Manuscript

768 Goudet J, Perrin N, Waser P (2002) Tests for sex-biased dispersal using bi-parentally inherited
769 genetic markers. *Molecular Ecology* **11**, 1103-1114.
770 Grant WS (2015) Problems and cautions with sequence mismatch analysis and Bayesian skyline
771 plots to infer historical demography. *Journal of Heredity* **106**, 333-346.
772 Hamner RM (2014) *All in a DNA's work: conservation genetics and monitoring of the New*
773 *Zealand endemic Maui's and Hector's dolphins.* PhD dissertation, Oregon State
774 University.
775 Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial
776 DNA mismatch distribution. *Human Biology* **66**, 591-600.
777 Hedrick PW, Allendorf FW, Baker CS (2013) Estimation of male gene flow from measures of
778 nuclear and female genetic differentiation. *Journal of Heredity* **104**, 713-717.
779 Ho SYW, Lanfear R, Bromham L*, et al.* (2011a) Time-dependent rates of molecular evolution.
780 *Molecular Ecology* **20**, 3087-3101.
781 Ho SYW, Lanfear R, Phillips MJ*, et al.* (2011b) Bayesian estimation of substitution rates from
782 ancient DNA sequences with low information content. *Systematic Biology* **60**, 366-375.
783 Hoelzel AR, Hey J, Dahlheim ME*, et al.* (2007) Evolution of population structure in a highly
784 social top predator, the killer whale. *Molecular Biology and Evolution* **24**, 1407-1415.
785 Ibáñez CM, Argüelles J, Yamashiro C*, et al.* (2012) Spatial genetic structure and demographic
786 inference of the Patagonian squid *Doryteuthis gahi* in the south-eastern Pacific Ocean.
787 *Journal of the Marine Biological Association of the United Kingdom* **92**, 197-203.
788 Ibáñez CM, Cubillos LA, Tafur R*, et al.* (2011) Genetic diversity and demographic history of
789 *Dosidicus gigas* (Cephalopoda: Ommastrephidae) in the Humboldt Current System. *Mar*
790 *Ecol Prog Ser* **431**, 163-171.
791 Ibáñez CM, Poulin E (2014) Genetic structure and diversity of squids with contrasting life
792 histories in the Humboldt Current system. *Hidrobiológica* **24**, 1-10.
793 Ivashchenko YV, Clapham PJ, Brownell Jr. RL (2013) Soviet catches of whales in the North
794 Pacific: revised totals. *Journal of Cetacean Research and Management* **13**, 59-71.
795 Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for
796 dealing with label switching and multimodality in analysis of population structure.
797 *Bioinformatics* **23**, 1801-1806.

This article is protected by copyright. All rights reserved

Author Manuscript

798 Janik VM (2001) Is cetacean social learning unique? *Behavioral and Brain Sciences* **24**, 337-338.
799 Jaquet N, Gendron D, Coakes A (2003) Sperm whales in the Gulf of California: residency,
800 movements, behavior, and the possible influence of variation in food supply. *Marine*
801 *Mammal Science* **19**, 545-562.
802 Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS
803 accommodates genotyping error increases success in paternity assignment. *Molecular*
804 *Ecology* **16**, 1099-1006.
805 Lambeck K, Rouby H, Purcell A, Sun Y, Sambridge M (2014) Sea level and global ice volumes
806 from the Last Glacial Maximum to the Holocene. *Proceedings of the National Academy*
807 *of Sciences* **111**, 15296-15303.
808 Lewis PO, Zaykin D (2001) Genetic Data Analysis: Computer program for the analysis of allelic
809 data v 1.0, available from http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php.
810 Lowther AD, Harcourt RG, Goldsworthy SD, Stow A (2012) Population structure of adult female
811 Australian sea lions is driven by fine-scale foraging site fidelity. *Animal Behaviour* **83**,
812 691-701.
813 Lyrholm T, Gyllensten U (1998) Global matrilineal population structure in sperm whales as
814 indicated by mitochondrial DNA sequences. *Proceedings of the Royal Society -*
815 *Biological Sciences (Series B)* **265**, 1679-1684.
816 Lyrholm T, Leimar O, Gyllensten U (1996) Low diversity and biased substitution patterns in the
817 mitochondrial DNA control region of sperm whales: implications for estimates of time
818 since common ancestry. *Molecular Biology and Evolution* **13**, 1318-1326.
819 Lyrholm T, Leimar O, Johanneson B, Gyllensten U (1999) Sex-biased dispersal in sperm whales:
820 contrasting mitochondrial and nuclear genetic structure of global populations.
821 *Proceedings of the Royal Society - Biological Sciences (Series B)* **266**, 347-347.
822 Martien KK, Baird RW, Hedrick NM*, et al.* (2012) Population structure of island-associated
823 dolphins: evidence from mitochondrial and microsatellite markers for common bottlenose
824 dolphins (*Tursiops truncatus*) around the main Hawaiian Islands. *Mar Mammal Sci* **28**,
825 E208-E232.

Author Manuscript

This article is protected by copyright. All rights reserved

826 Mate BR, Stafford KM, Ljungblad DK (1994) A change in sperm whale (*Physeter*
827 *macrocephalus*) distribution correlated to seismic surveys in the Gulf of Mexico. *The*
828 *Journal of the Acoustical Society of America* **96**, 3268-3269.
829 Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures.
830 *Molecular Ecology Resources* **11**, 5-18.
831 Meirmans PG, van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the
832 analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.
833 Mesnick SL, Engelhaupt D, Lyrholm T, Taylor BL, Escorza-Trevino S (2005) *Population*
834 *structure and the social animal: global genetic differentiation in sperm whales, Physeter*
835 *macrocephalus. Cachalot Assessment Research Planning (CARP) Workshop*.
836 Mesnick SL, Taylor BL, Archer FI, *et al.* (2011) Sperm whale population structure in the eastern
837 and central North Pacific inferred by the use of single-nucleotide polymorphisms,
838 microsatellites and mitochondrial DNA. *Molecular Ecology Resources* **11**, 278-298.
839 Mizroch SA, Rice DW (2013) Ocean nomads: distribution and movements of sperm whales in
840 the North Pacific shown by whaling data and Discovery marks. *Marine Mammal Science*
841 **29**, E136-E165.
842 Moors-Murphy HB (2014) Submarine canyons as important habitat for cetaceans, with special
843 reference to the Gully: A review. *Deep Sea Research Part II: Topical Studies in*
844 *Oceanography* **104**, 6-19.
845 Morin PA, Martien KK, Archer FI, *et al.* (2010) Applied conservation genetics and the need for
846 quality control and reporting of genetic data used in fisheries and wildlife management.
847 *Journal of Heredity* **101**, 1-10.
848 Morin PA, Parsons KM, Archer FI, *et al.* (2015) Geographic and temporal dynamics of a global
849 radiation and diversification in the killer whale. *Molecular Ecology* **24**, 3964-3979.
850 Moura AE, Janse van Rensburg C, Pilot M, *et al.* (2014) Killer Whale Nuclear Genome and
851 mtDNA Reveal Widespread Population Bottleneck during the Last Glacial Maximum.
852 *Molecular Biology and Evolution* **31**, 1121-1131.
853 Notarbartolo-Di-Sciara G (2014) Sperm whales, *Physeter macrocephalus*, in the Mediterranean
854 Sea: a summary of status, threats, and conservation recommendations. *Aquatic*
855 *Conservation: Marine and Freshwater Ecosystems* **24**, 4-10.

This article is protected by copyright. All rights reserved

Author Manuscript

856 Oremus M, Gales R, Dalebout ML*, et al.* (2009) Worldwide mitochondrial DNA diversity and
857 phylogeography of pilot whales (*Globicephala* spp.). *Biol J Linn Soc* **98**, 729-744.
858 Oremus M, Poole MM, Albertson GR, Baker CS (2012) Pelagic or insular? Genetic
859 differentiation of rough-toothed dolphins in the Society Islands, French Polynesia.
860 *Journal of Experimental Marine Biology and Ecology* **432-433**, 37-46.
861 Oremus M, Poole MM, Steel D, Baker CS (2007) Isolation and interchange among insular
862 spinner dolphin communities in the South Pacific revealed by individual identification
863 and genetic diversity. *Marine Ecology Progress Series* **336**, 275-289.
864 Ortega-Ortiz JG, Engelhaupt D, Winsor M, Mate BR, Hoelzel AR (2012) Kinship of long-term
865 associates in the highly social sperm whale. *Molecular Ecology* **21**, 732-744.
866 Ottensmeyer CA, Whitehead H (2003) Behavioural evidence for social units in long-finned pilot
867 whales. *Canadian Journal of Zoology* **81**, 1327-1338.
868 Palsbøll PJ, Bérubé M, Larsen AH, Jørgensen H (1997) Primers for the amplification of tri-and
869 tetramer microsatellite loci in baleen whales. *Molecular Ecology* **6**, 893-895.
870 Pannell JR, Whitlock M (2003) Coalescence in a metapopulation with recurrent local extinction
871 and recolonization. *Evolution* **57**, 949-961.
872 Pardini AT, Jones CS, Noble LR*, et al.* (2001) Sex-biased dispersal of great white sharks. *Nature*
873 **412**, 139-140.
874 Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic
875 software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.
876 Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic
877 software for teaching and research - an update. *Bioinformatics* **28**, 2537-2539.
878 Petit R (2010) PERMUT v 2.0. Available from
879 http://www.pierroton.inra.fr/genetics/labo/Software/PermutCpSSR/index.html.
880 Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences
881 and solutions. *Nature Reviews Genetics* **6**, 847-859.
882 Pons O, Petit RJ (1996) Measuring and Testing Genetic Differentiation With Ordered Versus
883 Unordered Alleles. *Genetics* **144**, 1237-1245.
884 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus
885 genotype data. *Genetics* **155**, 945-959.

This article is protected by copyright. All rights reserved

Author Manuscript

886 Ptak SE, Przeworski M (2002) Evidence for population growth in humans is confounded by fine-
887 scale population structure. *Trends in Genetics* **18**, 559-563.
888 R Core Team (2013) R: A language and environment for statistical computing. URL
889 http://www.R-project.org/. R Foundation for Statistical Computing, Vienna, Austria.
890 Raymond M, Rousset F (1995a) An exact test for population differentiation. *Evolution* **49**, 1280-
891 1283.
892 Raymond M, Rousset F (1995b) GENEPOP (Version 1.2): Population genetics software for exact
893 tests and ecumenicism. *Journal of Heredity* **86**, 248-249.
894 Rendell L, Mesnick S, Dalebout M, Burtenshaw J, Whitehead H (2012) Can genetic differences
895 explain vocal dialect variation in sperm whales, *Physeter macrocephalus? Behavior*
896 *Genetics* **42**, 332-343.
897 Richard KR, Dillon MC, Whitehead H, Wright JM (1996a) Patterns of kinship in groups of free-
898 living sperm whales (*Physeter macrocephalus*) revealed by multiple molecular genetic
899 analyses. *Proceedings of the National Academy of Science, United States of America* **93**,
900 8792-8795.
901 Richard KR, McCarrey SW, Wright JM (1994) DNA sequence from the *SRY* gene of the sperm
902 whale (*Physeter macrocephalus*) for use in molecular sexing. *Canadian Journal of*
903 *Zoology* **72**, 873-877.
904 Richard KR, Whitehead H, Wright JM (1996b) Polymorphic microsatellites from sperm whales
905 and their use in the genetic identification of individuals from naturally sloughed pieces of
906 skin. *Molecular Ecology* **5**, 313-315.
907 Riesch R, Barrett-Lennard LG, Ellis GM, Ford JKB, Deecke VB (2012) Cultural traditions and
908 the evolution of reproductive isolation: ecological speciation in killer whales? *Biological*
909 *Journal of the Linnean Society* **106**, 1-17.
910 Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise
911 genetic differences. *Molecular Biology and Evolution* **9**, 552-569.
912 Rokas A, Atkinson RJ, Brown GS, West SA, Stone GN (2001) Understanding patterns of genetic
913 diversity in the oak gallwasp *Biorhiza pallida*: demographic history or a *Wolbachia*
914 selective sweep? *Heredity* **87**, 294-304.

Author Manuscript

This article is protected by copyright. All rights reserved

915 Rousset F (2008) GENEPOP'007: a complete re-implementation of the genepop software for
916 Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
917 Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*, 2nd edn.
918 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
919 Savery LC, Evers DC, Wise SS*, et al.* (2013a) Global mercury and selenium concentrations in
920 skin from free-ranging sperm whales (*Physeter macrocephalus*). *Science of the Total*
921 *Environment* **450-451**, 59-71.
922 Savery LC, Wise SS, Falank C*, et al.* (2013b) Global assessment of silver pollution using sperm
923 whales (*Physeter macrocephalus*) as an indicator species. *Environmental & Analytical*
924 *Toxicology* **3**, 169.
925 Schlötterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in
926 cetacean species. *Nature* **354**, 63-65.
927 Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution
928 of pairwise differences when the mutation rates vary among sites: application to human
929 mitochondrial DNA. *Genetics* **152**, 1079-1089.
930 Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable
931 and exponentially growing populations. *Genetics* **129**, 555-562.
932 Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control
933 region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and*
934 *Evolution* **10**, 512-526.
935 Thompson KF, Patel S, Baker CS, Constantine R, Millar CD (2016) Bucking the trend: genetic
936 analysis reveals high diversity, large population size and low differentiation in a deep
937 ocean cetacean. *Heredity* **116**, 277-285.
938 Valsecchi E, Amos W (1996) Microsatellite markers for the study of cetacean populations.
939 *Molecular Ecology* **5**, 151-156.
940 van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software
941 for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology*
942 *Notes* **4**, 535-538.
943 Waldick RC, Brown MW, White BN (1999) Characterization and isolation of microsatellite loci
944 from the endangered North Atlantic right whale. *Molecular Ecology* **8**, 1763.

Author Manuscript

This article is protected by copyright. All rights reserved

945 Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic
946 methods for identifying the number of gene pools and their degree of connectivity.
947 *Molecular Ecology* **15**, 1419-1439.
948 Watkins WA, Schevill WE (1977) Sperm whale codas. *The Journal of the Acoustical Society of*
949 *America* **62**, 1485-1490.
950 Whitehead H (1993) The behaviour of mature male sperm whales on the Galápagos Islands
951 breeding grounds. *Canadian Journal of Zoology* **71**, 689-699.
952 Whitehead H (1994) Delayed competitive breeding in roving males. *Journal of Theoretical*
953 *Biology* **166**, 127-133.
954 Whitehead H (1998) Cultural selection and genetic diversity in matrilineal whales. *Science* **282**,
955 1708-1711.
956 Whitehead H (2002) Estimates of the current global population size and historical trajectory for
957 sperm whales. *Marine Ecology Progress Series* **242**, 295-304.
958 Whitehead H (2003) *Sperm whales: social evolution in the ocean* The University of Chicago
959 Press, Chicago.
960 Whitehead H (2005) Genetic diversity in the matrilineal whales: models of cultural hitchhiking
961 and group-specific non-heritable demographic variation. *Mar Mammal Sci* **21**, 58-79.
962 Whitehead H, Antunes R, Gero S*, et al.* (2012) Multilevel societies of female sperm whales
963 (*Physeter macrocephalus*) in the Atlantic and Pacific: why are they so different?
964 *International Journal of Primatology* **33**, 1142-1164.
965 Whitehead H, Christal J, Dufault S (1997) Past and distant whaling and the rapid decline of
966 sperm whales off the Galápagos Islands. *Conservation Biology* **11**, 1387-1396.
967 Whitehead H, Coakes A, Jaquet N, Lusseau S (2008) Movements of sperm whales in the tropical
968 Pacific. *Marine Ecology Progress Series* **361**, 291-300.
969 Whitehead H, Dillon M, Dufault S, Weilgart L, Wright J (1998) Non-geographically based
970 population structure of South Pacific sperm whales: dialects, fluke-markings and genetics.
971 *Journal of Animal Ecology* **67**, 253-262.
972 Winkelmann I, Campos PF, Strugnell J*, et al.* (2013) Mitochondrial genome diversity and
973 population structure of the giant squid *Architeuthis*: genetics sheds new light on one of the

Author Manuscript

This article is protected by copyright. All rights reserved

974 most enigmatic marine species. *Proceedings of the Royal Society B: Biological Sciences*
975 **280**, 20130273.
976 Wise Sr JP, Payne R, Wise SS*, et al.* (2009) A global assessment of chromium pollution using
977 sperm whales (*Physeter macrocephalus*) as an indicator species. *Chemosphere* **75**, 1461-
978 1467.
979 Wise Sr. JP, Thompson WD, Wise SS*, et al.* (2011) A global assessment of gold, titanium,
980 strontium and barium pollution using sperm whales (*Physeter macrocephalus*) as an
981 indicator species. *Journal of Ecosystem & Ecography* **1**, 101.
982
983
984

985 **8. Data accessibility statement**

986 -- For each sample genotyped in this study, general location (as well as latitude/longitude where
987 available), individual ID code, social group code, sex, microsatellite genotype, and a letter code
988 denoting the mtDNA CR haplotype (defined in Supplementary Material 1) are archived with
989 Dryad (doi:10.5061/dryad.2q4r0).
990 -- Sequences of each defined haplotype have been archived on NCBI GenBank (accession
991 numbers: KU719571 - KU719622).
992 -- Scripts used in analyses are available with online Supplementary Materials or at
993 https://github.com/laninsky/genetic_diversity_diffs
994
995

996 **9. Author contributions**

997 A. A. performed genetic analyses, wrote analysis scripts, analyzed data, and wrote the paper. D.S.
998 and K.H. performed genetic analyses and analyzed data. S.M. and D.E. provided analysis
999 recommendations. I.K. and R.P. designed the *Odyssey* sample collection strategy, performed field
1000 research and collected field data. C.S.B. supervised the research design and provided analysis
1001 recommendations. All authors provided editorial input to writing of the paper.

Author Manuscript

This article is protected by copyright. All rights reserved

1002
1003
1004
1005
1006
1007

1008 **Table 1:** Regional, oceanic and worldwide sample sizes (*n*) and diversity metrics for mtDNA and microsatellites. For mtDNA,
1009 number of haplotypes (*k*), haplotype diversity (*h*) and nucleotide diversity (in %, π) is presented, with standard deviations
1010 calculated in *Arlequin*. For microsatellites, numbers of individuals are given by sex (*F, M*) and total sample size (*n*). A binomial
1011 exact test was used to identify areas with a significant bias of females (asterisk after female sample size) or males (asterisk after
1012 male sample size), where * significant at $p < 0.05$; ** significant at $p < 0.001$. F_{IS} values are indicated as significant where *
1013 significant at $p < 0.05$. Regional allelic richness is adjusted by minimum regional sample size, with oceanic allelic richness
1014 adjusted by minimum ocean sample size. Regions ordered from east to west. 'Unassigned' includes samples not originating from
1015 tropical/subtropical regions, from areas with samples sizes too small to include in regional analyses, or those samples without a
1016 specific ~500 km regional location. References for data: [1] This study; [2] Lyrholm & Gyllensten (1998); [3] Rendell *et al.*
1017 (2012); [4] Richard *et al.* (1996a); [5] Whitehead *et al.* (1998); [6] Mesnick *et al.* (2011); [7] Engelhaupt *et al.* (2009).

1018

1019 **Table 1** (Continued)

1020

	mtDNA	microsatellites

Author Manuscript

This article is protected by copyright. All rights reserved

Author Manuscript

	Geographic area	*n*	*k*	*h (s. d.)*	π (*s.d.*)	*F*	*M*	*n*	*Allelic richness*	H_O	F_{IS}	*References*
Pacific	*Gulf of California*	122	11	0.788 (*0.024*)	0.368 (*0.250*)	93**	20	122	5.2	0.702	0.016	1
	Galapagos/Ecuador	285	16	0.744 (*0.012*)	0.352 (*0.240*)	0	23**	23	5.1	0.677	0.031	1, 2, 3, 4, 5
	Pacific Crossing	36	8	0.679 (*0.071*)	0.301 (*0.220*)	20	13	37	5.1	0.704	0.013	1
	Hawai'i	28	4	0.643 (*0.068*)	0.195 (*0.164*)	--	--	--	--	--	--	6
	Kiribati	13	4	0.718 (*0.089*)	0.381 (*0.276*)	10*	2	13	5.3	0.684	0.092*	1
	Papua New Guinea	63	8	0.720 (*0.036*)	0.299 (*0.216*)	54**	8	65	5.1	0.687	0.031	1
	Unassigned Pacific	478	--	--	--	22	31	66	**--**	**--**	**--**	1, 2, 3, 5, 6
	Total	***1025***	***33***	***0.780 (0.008)***	***0.385 (0.256)***	***199*******	***97***	***326***	***8.9***	***0.704***	***0.021****	**--**
Indian	*Southwestern Australia*	21	5	0.791 (*0.044*)	0.305 (*0.226*)	9	4	21	5.1	0.697	0.007	1
	Cocos Island	18	3	0.451 (*0.117*)	0.229 (*0.187*)	18**	0	18	5.2	0.712	0.001	1
	Sri Lanka	42	3	0.382 (*0.076*)	0.131 (*0.125*)	42**	6	56	5.1	0.671	0.040*	1
	Maldives/ Chagos Archipelago	33	4	0.570 (*0.061*)	0.300 (*0.220*)	9	15	34	5.3	0.700	0.041	1
	Seychelles	31	6	0.716 (*0.066*)	0.407 (*0.276*)	17**	2	31	5.3	0.697	0.020	1
	Aldabras	12	3	0.712 (*0.069*)	0.362 (*0.267*)	6	2	12	5.1	0.677	0.027	1
	Unassigned Indian	2	--	--	--	1	1	3	**--**	**--**		1
	Total	***159***	***8***	***0.788 (0.015)***	***0.426 (0.280)***	***102*******	***30***	***175***	***8.8***	***0.686***	***0.035****	**--**
Atlantic	*Mediterranean*	40	1	0.000 (*0.000*)	0.000 (*0.000*)	1	8*	9	4.6	0.631	0.086	1, 7
	Canary Islands	14	3	0.648 (*0.081*)	0.329 (*0.246*)	14	8	25	5.2	0.690	0.014	1
	Western North Atlantic	87	6	0.616 (*0.028*)	0.271 (*0.200*)	--	--	--	--	--	--	1, 7
	Gulf of Mexico	153	5	0.500 (*0.044*)	0.211 (*0.167*)	--	--	--	--	--	--	7
	Unassigned Atlantic	68	--	--	--	3	0	7	**--**	**--**	**--**	1, 2
	Total	***362***	***8***	***0.748 (0.010)***	***0.333 (0.231)***	***18***	***16***	***41***	***8.8***	***0.669***	***0.051****	**--**
	Unassigned Worldwide	41	**--**	**--**	**--**	--	**--**	**--**	**--**	**--**	**--**	2

This article is protected by copyright. All rights reserved

Worldwide total	1587	39	0.818 (0.005)	0.430 (0.279)	319**	143	542	14.0	0.696	0.029	--

1021
1022

1023 **Table 2**: Summary of microsatellite locus-specific characteristics for the 542 individuals genotyped in this study. *n* gives the
1024 number of individuals successfully typed at each locus. *Ho* and *He* (observed and expected heterozygosity, respectively)
1025 calculated in *Cervus*. F_{ST} calculated in *Genepop* for oceanic and regional subsets of data (see Table 1). Statistically significant
1026 F_{ST} values are bolded and italicized, with * significant at $p < 0.05$; ** significant at $p < 0.001$.
1027

Locus	*n*	Reference	Size range (bp)	No of alleles	*Ho*	*He*	Oceanic F_{ST}	Regional F_{ST}
EV1	521	Valsecchi & Amos (1996)	118 – 142	12	0.599	0.641	0.0034	0.0061
EV5	529	Valsecchi & Amos (1996)	148 – 174	11	0.711	0.708	***0.0100****	0.0071
EV14	483	Valsecchi & Amos (1996)	121 - 155	14	0.687	0.716	0.0032	***0.0120****
EV37	504	Valsecchi & Amos (1996)	177 - 250	32	0.855	0.905	***0.0029****	***0.0050****
EV94	534	Valsecchi & Amos (1996)	193 - 225	17	0.82	0.804	0.0017	0.0025
GATA417	438	Palsbøll *et al.* (1997)	172 - 202	7	0.509	0.532	0.0107	0.0019
GT23	523	Bérubé *et al.* (2000)	75 - 99	12	0.511	0.499	0.0034	0.0000
GT575	487	Bérubé *et al.* (2000)	131 - 137	4	0.61	0.611	0.0011	0.0104
rw4-10	461	Waldick *et al.* (1999)	177 - 213	14	0.72	0.768	***0.0028****	0.0037
SW13	523	Richard *et al.* (1996b)	134 - 176	14	0.824	0.835	0.0000	***0.0092****
464/465	404	Schlötterer *et al.* (1991)	141 - 145	3	0.527	0.541	0.0035	0.0000

This article is protected by copyright. All rights reserved

Author Manuscript

SW19	508	Richard *et al.* (1996b)	89 - 167	32	0.88	0.921	***0.0017****	***0.0029*****
FCB1	519	Buchanan *et al.* (1996)	107 - 145	16	0.792	0.835	0.0018	0.0032
Average	494.9			14.5	0.696	0.717	***0.0032*****	***0.0048*****

1028

Author Manuscript

This article is protected by copyright. All rights reserved

Author Manuscript

Table 3: Degrees of freedom (*d. f.*) and percentage of variation (%) explained by ocean, region, social group, and among individuals for hierarchical AMOVAs, nesting social group within ocean for the worldwide dataset, and social group within region for the worldwide, Pacific and Indian Ocean datasets. (a) mtDNA F_{ST}; (b) mtDNA Φ_{ST}; and (c) microsatellite F_{ST}. Levels which explain a significant percentage of variation are bolded and italicized, with * significant at $p < 0.05$; ** significant at $p < 0.001$. The social groups these results are based on are summarized in Supplementary Material 5.

(a)	Worldwide				Pacific		Indian	
mtDNA F_{ST}	*d.f.*	%	*d.f.*	%	*d.f.*	%	*d.f.*	%
Among oceans	2	***15.1***	--	--	--	--	--	--
Among regions	--	--	7	***22.7***	2	-3.0	3	***44.4***
Among social groups	31	***32.0***	22	***26.9***	9	***34.4***	6	***12.3***
Among individuals	350	***52.9***	292	***50.4***	162	***68.6***	90	***43.2****

(b)	Worldwide				Pacific		Indian	
mtDNA Φ_{ST}	*d.f.*	%	*d.f.*	%	*d.f.*	%	*d.f.*	%
Among oceans	2	***25.0***	--	--	--	--	--	--
Among regions	--	--	7	***34.9***	2	-1.3	3	***51.9***
Among social groups	31	***29.8***	22	***21.5***	9	***31.9***	6	***10.6****
Among individuals	350	***45.2***	292	***43.6***	162	***69.4***	90	***37.6****

(c)	Worldwide				Pacific		Indian	
microsatellite F_{ST}	*d.f.*	%	*d.f.*	%	*d.f.*	%	*d.f.*	%
Among oceans	2	0.27	--	--	--	--	--	--
Among regions	--	--	6	0.07	1	0.03	3	-0.25
Among social groups	23	***1.32***	15	***1.49***	7	***1.25***	7	***1.66***
Among individuals	281	98.4	222	98.4	110	98.7	104	98.6

Table 4: Sex-specific F_{ST} comparisons by marker and estimates of sex-biased gene flow (*Nm*, m_M/m_F, Hedrick *et al.*, 2013). Regional analyses of sex-specific F_{ST} were limited to

This article is protected by copyright. All rights reserved

1040 areas with more than two identified females and males as summarized in **Table 1**. Note,
1041 that although male-specific microsatellite F_{ST} appears to exceed that of females among
1042 regions in the Pacific, neither estimate is significantly different from zero. Due to limited
1043 sample sizes, a within-ocean regional F_{ST} analysis was not conducted for the Atlantic.
1044 Hedrick *et al.*'s (2013) estimates of sex-specific gene flow are based on the fixation indices
1045 presented in **Figure 4**. As all variance in assignment tests (vAIc) were not significant,
1046 results of these tests are not displayed. Statistically significant values (for the *p*-values for
1047 the difference in F_{ST} between sexes) are bolded and italicized, with * significant at $p <$
1048 0.05; ** significant at $p < 0.001$.
1049

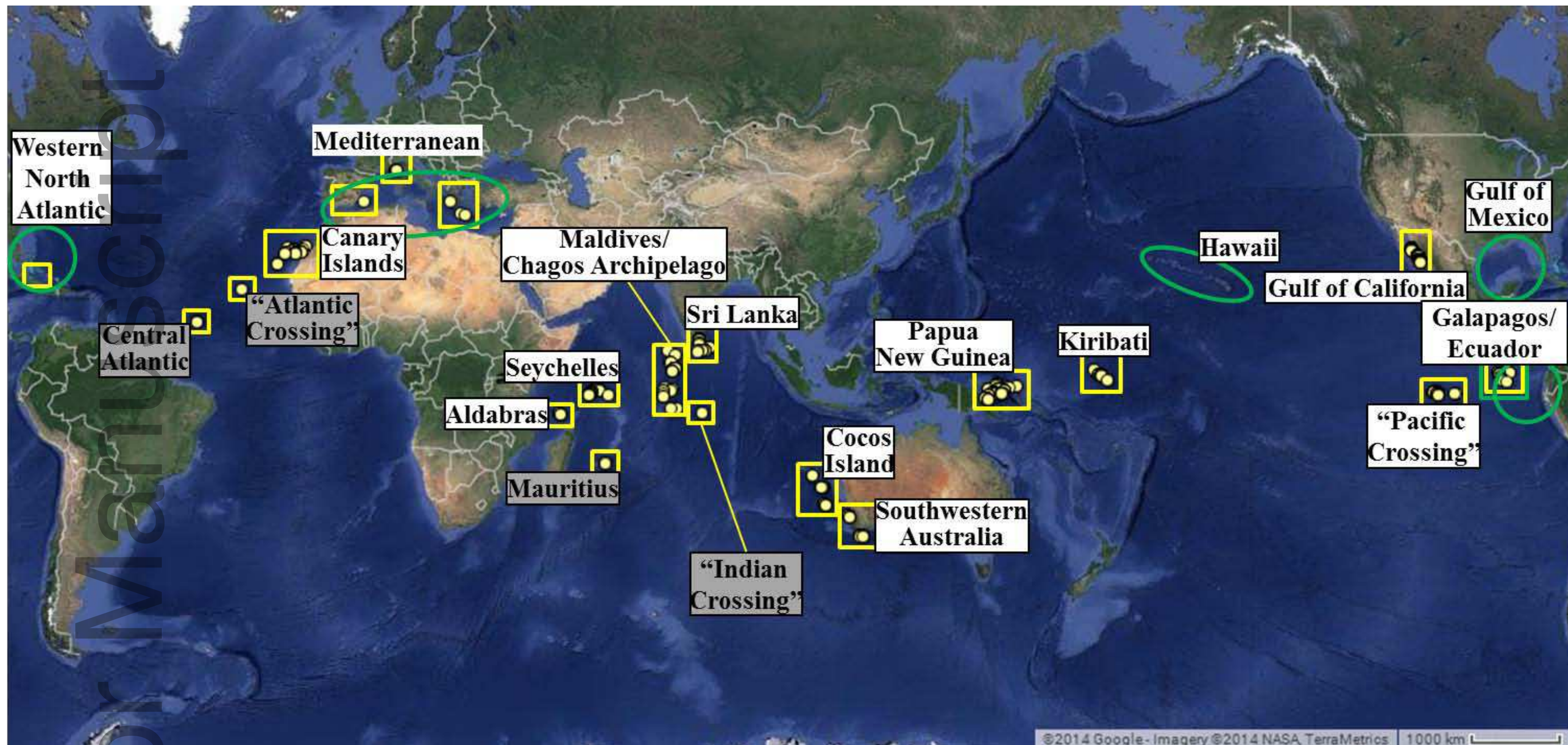
Area		Sex	mtDNA CR			Microsatellites			Gene flow	
			n	F_{ST}	*p* -value	n	F_{ST}	*p* -value	*Nm*	m_M/m_F
Pacific	By region	F	175	0.1145	0.1632	177	0.0004	0.8438	7.75	22.01
		M	42	0.0640		43	0.0103		170.57	
Indian	By region	F	70	0.4892	0.1666	83	0.0061	0.4499	1.09	43.56
		M	27	0.2878		29	0.0050		47.68	
Atlantic	By region	F	--	--	--	--	--	--	0.57	25.99
		M							14.71	
Worldwide	By region	F	253	0.2735	***0.0366****	274	0.0063	0.2745	1.45	34.82
		M	72	0.1426		80	0.0028		50.39	
	By ocean	F	289	0.1259	0.0725	319	0.0068	***0.0351****	4.38	16.77
		M	118	0.0673		143	0.0008		73.49	

1050
1051
1052
1053

Author Manuscript

This article is protected by copyright. All rights reserved

1054



1055

1056 **Figure 1:** Distribution of genetic samples from sperm whales used in analyses. White labels indicate areas included in within-ocean
1057 mtDNA and microsatellite tropical/subtropical regional analyses. *Odyssey* samples were aggregated together if they occurred within
1058 500 km of another sample. This created the localized regional areas shown in the rectangles. Additional mtDNA samples/regions
1059 included in analyses originating from previous studies collected over similar spatial scales are circled (references in Table 1). Gray
1060 labels show regions not included in regional analyses due to small sample sizes.

Author Manuscript

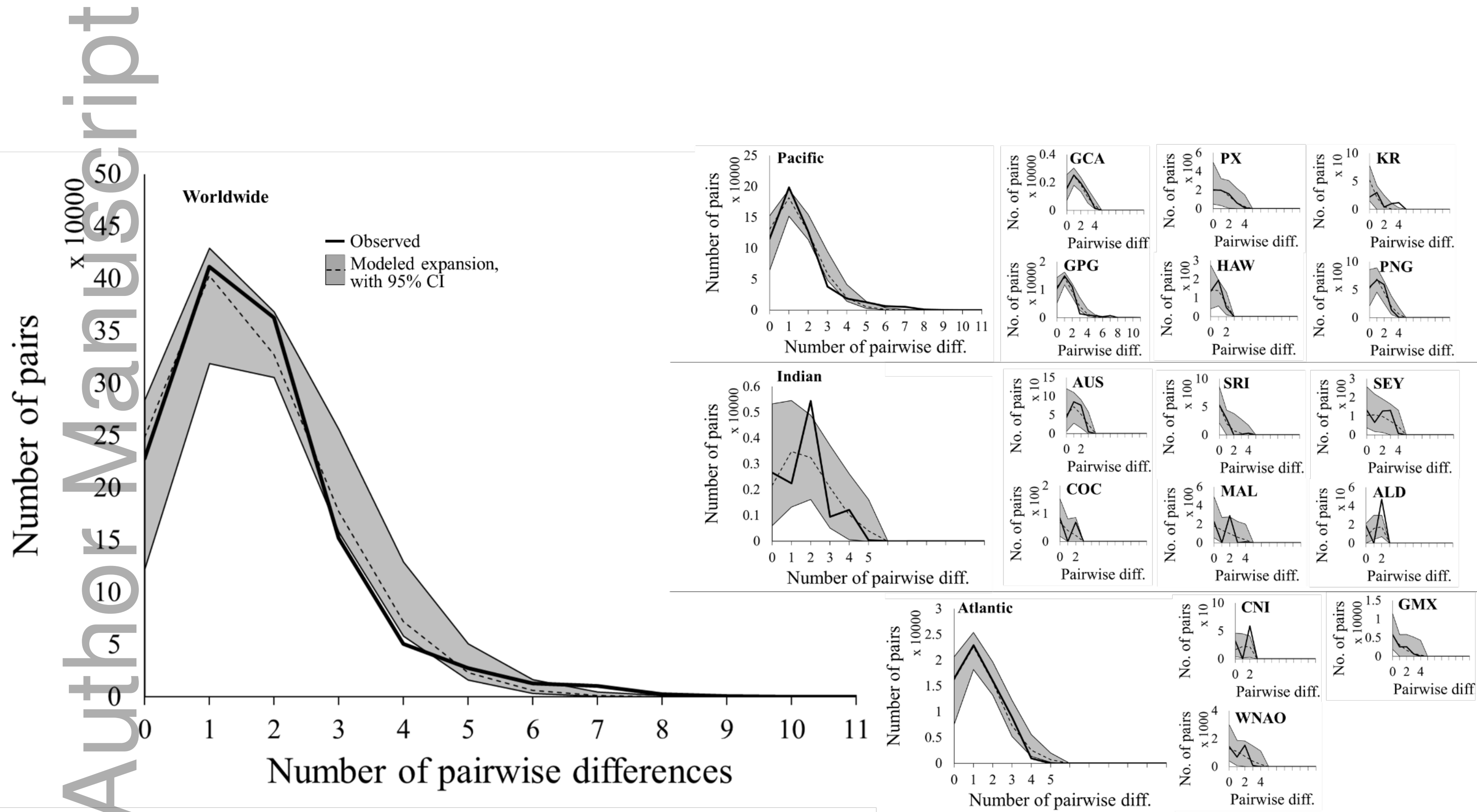
This article is protected by copyright. All rights reserved

1061

1062

1063

1064



Author Manuscript

1065

1066

This article is protected by copyright. All rights reserved

1067 **Figure 2:** Mismatch distributions for mtDNA at worldwide, oceanic and regional levels. Abbreviations: Gulf of California (GCA),
1068 Galapagos/Ecuador (GPG), 'Pacific Crossing' (PX), Hawai'i (HAW), Kiribati (KR), Papua New Guinea (PNG), SW Australia (AUS),
1069 Cocos Island (COC), Sri Lanka (SRI), Maldives/Chagos Archipelago (MAL), Seychelles (SEY), Aldabras (ALD), Canary Island
1070 (CNI), Western North Atlantic (WNAO), Gulf of Mexico (GMX). A mismatch distribution was not generated for the Mediterranean
1071 due to lack of mtDNA variation within this region.
1072
1073
1074
1075
1076
1077
1078 **Figure 3:** Maximum parsimony network based on 394 bp of the mtDNA CR (haplotype definitions in Supplementary Material 1).
1079 Haplotypes are colored by ocean, with the exception of haplotypes V and W that were not localized to a specific ocean in previous
1080 studies. Size of haplotype pie is proportional on a log scale to the total number of samples with the haplotype. Lines represent
1081 substitutions (one or two between haplotypes as defined by the key). New haplotypes characterized in this study are outlined.
1082 **Figure 3** (Continued)
1083

Author Manuscript

This article is protected by copyright. All rights reserved

1084
1085
1086
1087
1088

Author Manuscript

This article is protected by copyright. All rights reserved

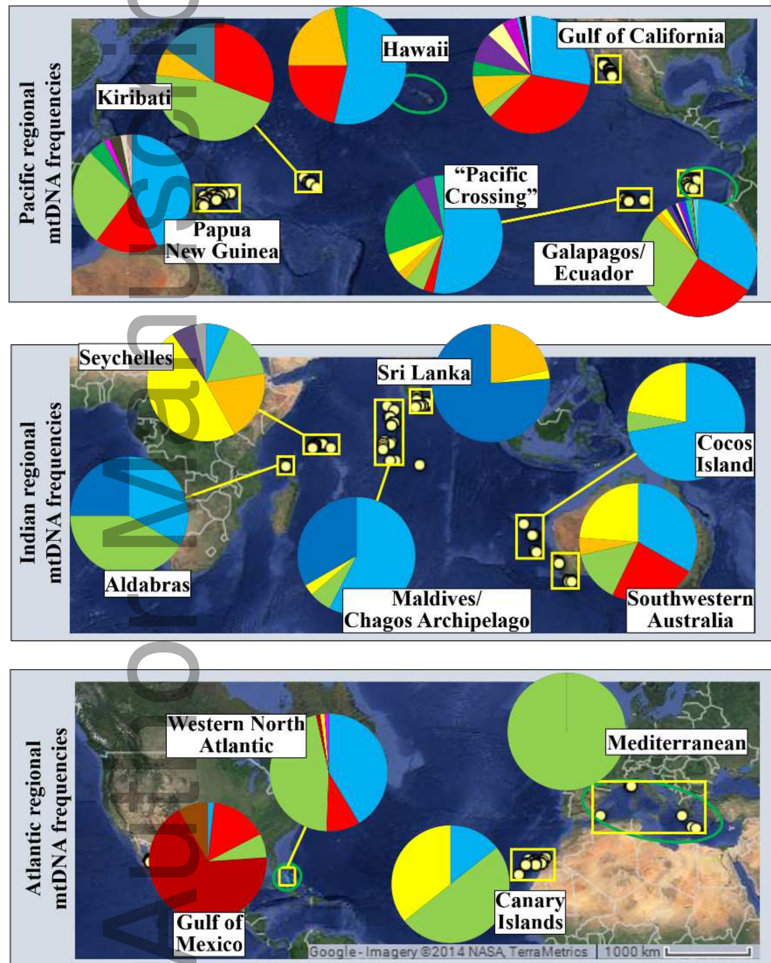
1089

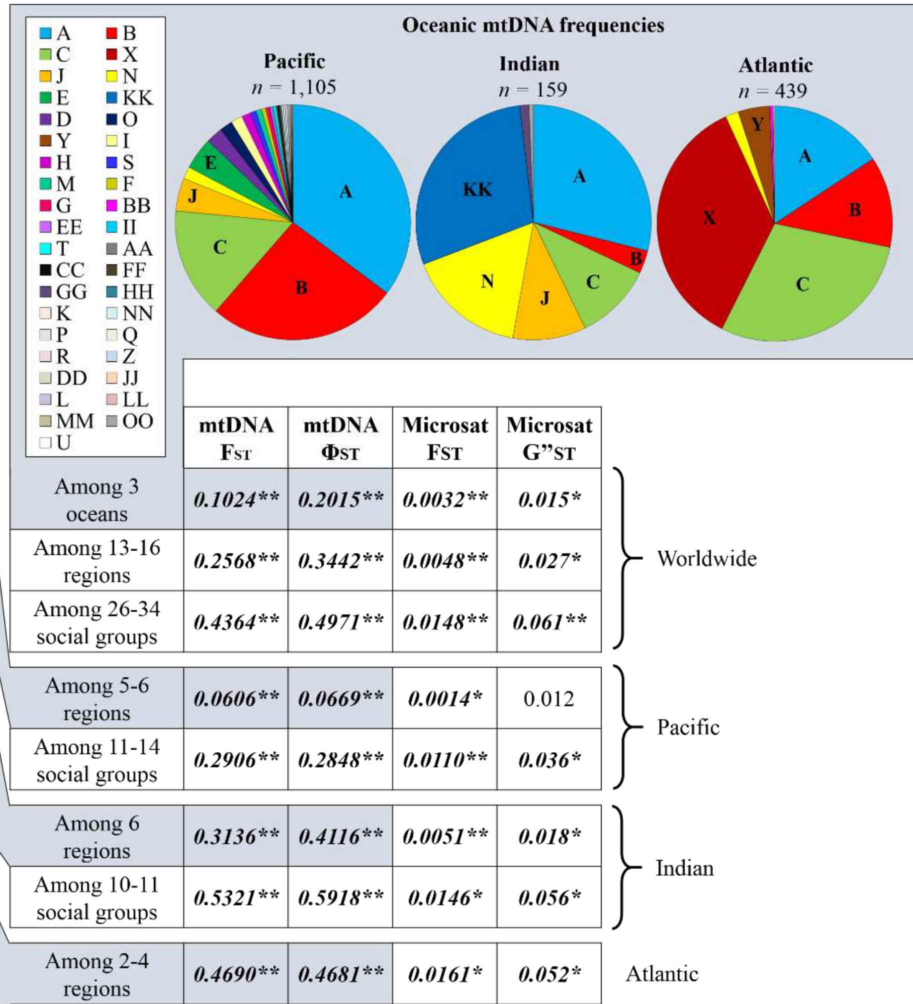
1090

1091 **Figure 4:** Differentiation of mtDNA and microsatellites at oceanic, regional and social group levels with regional/oceanic mtDNA
1092 haplotype frequencies. Where a range in number of social groups/regions are given, the smaller number corresponds to the
1093 microsatellite sample size and the larger number to the mtDNA sample size. No social group analysis was conducted for the Atlantic
1094 due to limited sample sizes. Braces to right of table give scale of each analysis (worldwide, and by each ocean). See Table 1 for sample
1095 sizes used in regional analyses and Supplementary Material 5 for sample sizes used in social group analyses. Pairwise comparisons at
1096 oceanic and regional levels are given in Supplementary Material 12. Haplotype key ordered by worldwide abundance of haplotype. *
1097 significant at $p < 0.05$; ** significant at $p < 0.001$.
1098 **Figure 4** (Continued)

Author Manuscript

This article is protected by copyright. All rights reserved





	mtDNA F_{ST}	mtDNA Φ_{ST}	Microsat F_{ST}	Microsat G''_{ST}	
Among 3 oceans	*0.1024***	*0.2015***	*0.0032***	*0.015**	Worldwide
Among 13-16 regions	*0.2568***	*0.3442***	*0.0048***	*0.027**	
Among 26-34 social groups	*0.4364***	*0.4971***	*0.0148***	*0.061***	
Among 5-6 regions	*0.0606***	*0.0669***	*0.0014**	0.012	Pacific
Among 11-14 social groups	*0.2906***	*0.2848***	*0.0110***	*0.036**	
Among 6 regions	*0.3136***	*0.4116***	*0.0051***	*0.018**	Indian
Among 10-11 social groups	*0.5321***	*0.5918***	*0.0146**	*0.056**	
Among 2-4 regions	*0.4690***	*0.4681***	*0.0161**	*0.052**	Atlantic

1099

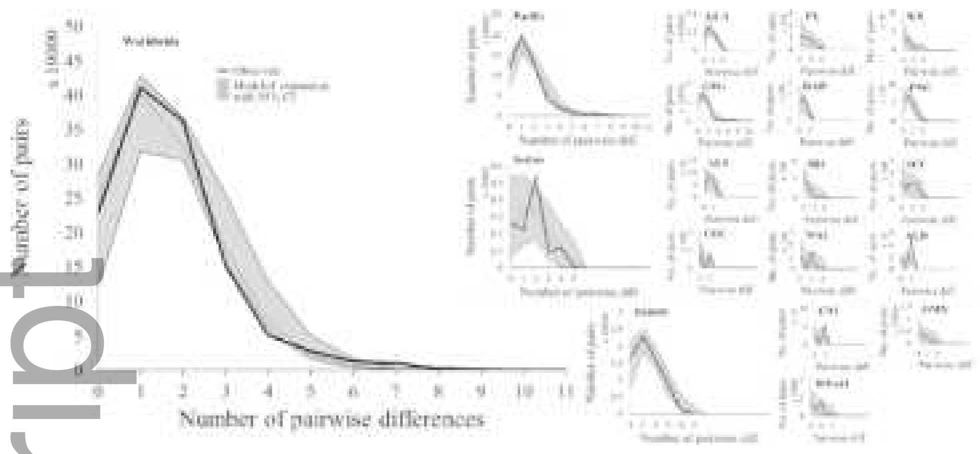
Author Manuscript

This article is protected by copyright. All rights reserved

mec_13638_f1.png

Author Manuscript

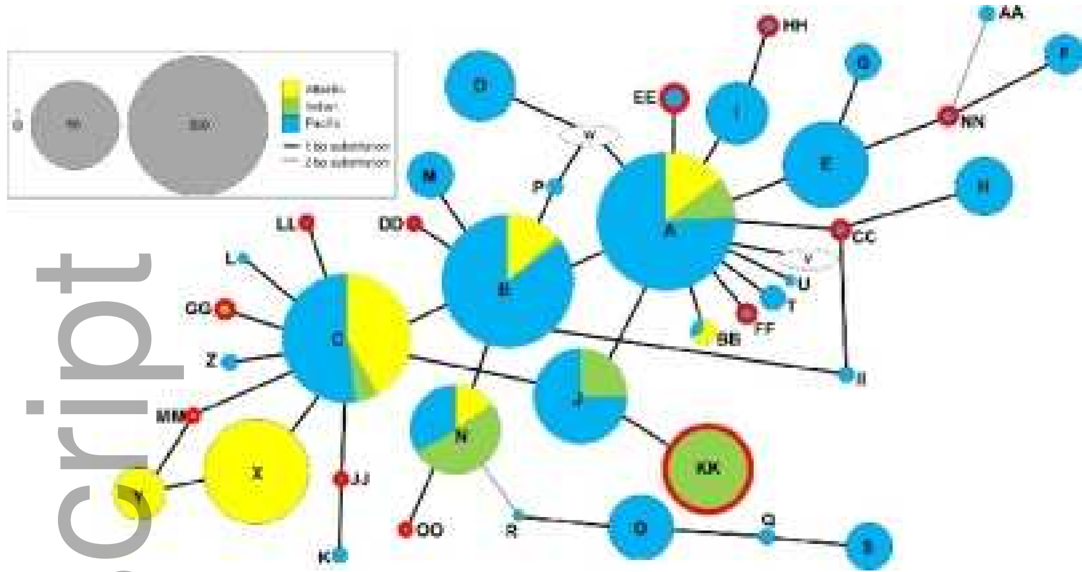
This article is protected by copyright. All rights reserved



mec_13638_f2.png

Author Manuscript

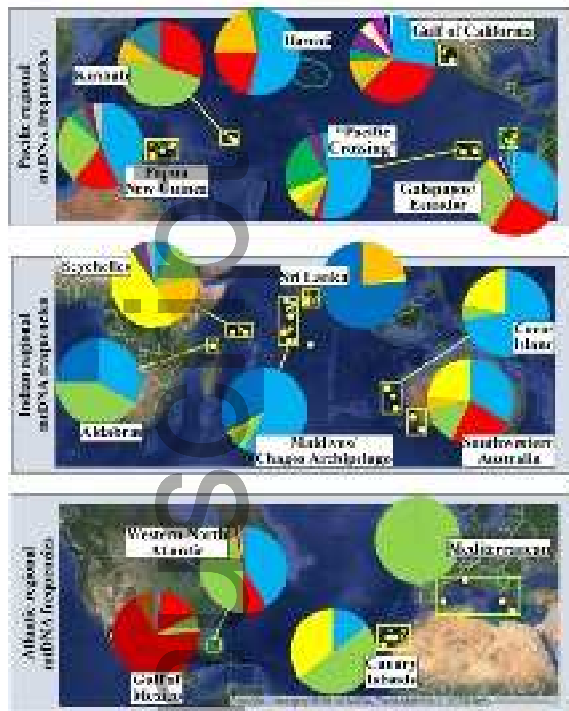
This article is protected by copyright. All rights reserved

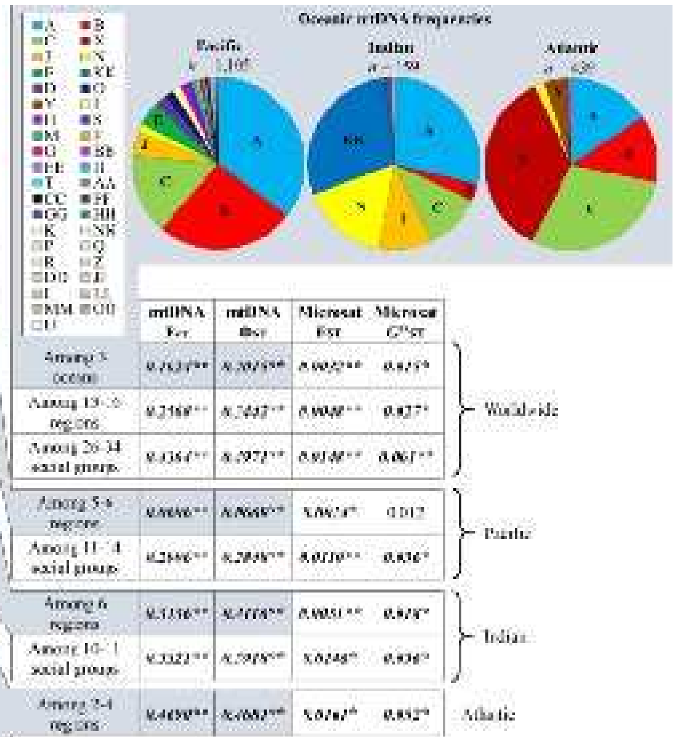


mec_13638_f3.png

Author Manuscript

This article is protected by copyright. All rights reserved





	mtDNA F_{ST}	mtDNA Φ_{ST}	Microsat F_{ST}	Microsat G''_{ST}	
Among 3 oceans	[illegible]	[illegible]	[illegible]	[illegible]	Worldwide
Among 13–16 regions	[illegible]	[illegible]	[illegible]	0.027*	Worldwide
Among 26–34 social groups	[illegible]	[illegible]	[illegible]	0.061**	Worldwide
Among 5–6 regions	[illegible]	[illegible]	[illegible]	0.012	Pacific
Among 11–14 social groups	[illegible]	[illegible]	[illegible]	0.036*	Pacific
Among 6 regions	[illegible]	[illegible]	[illegible]	[illegible]	Indian
Among 10–11 social groups	[illegible]	[illegible]	[illegible]	[illegible]	Indian
Among 3–4 regions	[illegible]	[illegible]	[illegible]	0.052*	Atlantic

mec_13638_f4.png

Author Manuscript

This article is protected by copyright. All rights reserved