

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/265912026>

# Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift

Article in *Molecular Ecology* · September 2014

DOI: 10.1111/mec.12929

CITATIONS

60

READS

471

10 authors, including:



**Andre Moura**

Muzeum i Instytut Zoologii PAN

51 PUBLICATIONS 805 CITATIONS

[SEE PROFILE](#)



**Margaret A Hughes**

University of Liverpool

98 PUBLICATIONS 3,177 CITATIONS

[SEE PROFILE](#)



**Andreanna J Welch**

Durham University

20 PUBLICATIONS 1,119 CITATIONS

[SEE PROFILE](#)



**Ryan R Reisinger**

University of Southampton

74 PUBLICATIONS 1,437 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Review of African cetacean research [View project](#)



The *Bactrocera oleae* gut microflora: function, interactions, evolution [View project](#)

Received Date : 06-May-2014

Revised Date : 09-Sep-2014

Accepted Date : 15-Sep-2014

Article type : Original Article

**Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift**

Andre E. Moura<sup>1†</sup>, John G. Kenny<sup>2</sup>, Roy Chaudhuri<sup>2</sup>, Margaret A. Hughes<sup>2</sup>, Andreanna Welch<sup>1</sup>, Ryan R. Reisinger<sup>3</sup>, P. J. Nico de Bruyn<sup>3</sup>, Marilyn E. Dahlheim<sup>4</sup>, Neil Hall<sup>2</sup>, A. Rus Hoelzel<sup>1</sup>

<sup>1</sup> School of Biological and Biomedical Sciences, Durham University, South Road, Durham, DH1 3LE, UK

<sup>2</sup> Department of Functional and Comparative Genomics, Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK

<sup>3</sup> Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Private Bag X20, Hatfield 0028, Pretoria, South Africa

<sup>4</sup> National Marine Mammal Laboratory, National Marine Fisheries Service, 7600 Sand Point Way N.E., Seattle, WA 98115, USA

This article has been accepted for publication and 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.12929

This article is protected by copyright. All rights reserved.

<sup>†</sup> Current Address: School of Life Sciences, University of Lincoln, Lincoln LN2 2LG, UK

Corresponding author: A. Rus Hoelzel. Address: School of Biological and Biomedical Sciences, Durham University, South Road, Durham, DH1 3LE, UK; Telephone: +44 (0) 191 33 41325. Email: a.r.hoelzel@dur.ac.uk

## Abstract

The evolution of diversity in the marine ecosystem is poorly understood, given the relatively high potential for connectivity, especially for highly mobile species such as whales and dolphins. The killer whale (*Orcinus orca*) has a worldwide distribution, and individual social groups travel over a wide geographic range. Even so, regional populations have been shown to be genetically differentiated, including among different foraging specialists (ecotypes) in sympatry. Given the strong matrifocal social structure of this species together with strong resource specialisations, understanding the process of differentiation will require an understanding of the relative importance of both genetic drift and local adaptation. Here we provide a high resolution analysis based on nuclear SNP markers and inference about differentiation at both neutral loci and those potentially under selection. We find that all population comparisons, within or among foraging ecotypes, show significant differentiation, including populations in parapatry and sympatry. Loci putatively under selection show a different pattern of structure compared to neutral loci, and are associated with gene ontology terms reflecting physiologically relevant functions (e.g. related to digestion). The pattern of differentiation for one ecotype in the North Pacific suggests local

Accepted Article

adaptation and shows some fixed differences among sympatric ecotypes. We suggest that differential habitat use and resource specialisations have promoted sufficient isolation to allow differential evolution at neutral and functional loci, but that the process is recent and dependent on both selection and drift.

## Introduction

The scarcity of strong geographical barriers and the large dispersal abilities of many marine species have led to suggestions that selection is a dominant mechanism driving population differentiation (Naciri et al. 1999; Bierne et al. 2003). This view has been supported by studies showing genetic structure associated with environmental and physiological differences (Ruzzante et al. 1996; Beheregaray and Sunnucks 2001; Jorgensen et al. 2005; Palumbi and Lessios 2005; Atarhouch et al. 2007; White et al. 2010; Bowen et al. 2013). However, several examples are known among marine taxa where small scale population differentiation has been attributed to complex phylogeographic scenarios involving differentiation in isolation due to cyclical climate change (Alvarado Bremer et al. 2005; Domingues et al. 2007; Domingues et al. 2008), vicariance, or the opening of dispersal corridors (Tringali 1999; de Bruyn et al. 2005; Puckridge et al. 2012).

Cetaceans have large dispersal abilities and long life expectancies, and yet often show genetic differentiation over relatively small geographic scales (Hoelzel 2009). Some studies have suggested that small scale population differentiation in these species is caused by adaptation to local environmental differences (Yoshida et al. 2001; Möller et al. 2010;

Mendez et al. 2011; Amaral et al. 2012), though the inference is indirect. It is also the case that closely related species with very similar distribution patterns may show very different patterns of genetic differentiation, such as the fine scale differentiation seen for bottlenose dolphins (*Tursiops truncatus*) in European waters (Natoli et al. 2005) compared to the lack of structure for common dolphins (*Delphinus delphis*) over the same range (Moura et al. 2013a). This may emphasize the importance of species-specific resource requirements and specialisations towards the evolution of habitat dependence, philopatry and population differentiation (Yurk et al. 2002; Natoli et al. 2006; Hoelzel et al. 2007; Moura et al. 2013a).

Killer whales show population genetic structure over spatial scales that are much smaller than their dispersal abilities (Hoelzel et al. 2007). These large-bodied dolphins are distributed worldwide and organized into stable, matrifocal social groups called pods. Different communities of pods exhibit consistent, long-term specialisations on prey resource, defining different 'ecotypes' (which sometimes also differ with respect to other aspects of behaviour and morphology, see de Bruyn et al. 2013). Although the level of gene flow between pods varies depending on the ecotype, gene flow between different ecotypes has been shown to be limited based on inference from both mtDNA and microsatellite DNA markers (Hoelzel et al. 2007; Morin et al. 2010), though there may be exceptions (Foote et al. 2013). In the North Pacific, two ecotypes, known as 'residents' and 'transients' occupy largely sympatric distribution ranges (Ford et al. 2000), but are genetically well differentiated (e.g. Hoelzel et al. 2007). This is coincident with differences in prey specialization (fish vs marine mammals respectively, Ford et al. 1998; Krahn et al. 2007),

Accepted Article

social organization (Ford et al. 2000), mating systems (Pilot et al. 2010), and vocal behaviour (Yurk et al. 2002; Deecke et al. 2005).

Significant genetic differentiation is typically found for all comparisons of killer whale populations defined *a priori* either geographically or by ecotype (Hoelzel et al. 2007; Morin et al. 2010), though overall diversity is low worldwide, likely due to a bottleneck during the last glacial period (Hoelzel et al. 2002; Moura et al. 2014a). Differentiation is seen both between ecotypes in sympatry and following a pattern of isolation by distance within an ecotype (Hoelzel et al. 2007). However, previous studies restricted to neutral markers can provide only limited insight into the mechanisms of ecological adaptation and differentiation between ecotypes. Here we focus on the North Pacific, but include outgroup populations from the North Atlantic (Iceland) and Southern Oceans (Marion Island). We use restriction-site associated DNA (RAD) single nucleotide polymorphic (SNP) markers to provide a high resolution genome-wide assessment of population structure at both neutral loci and markers putatively under selection. We test the hypothesis that populations representing sympatric ecotypes (e.g. residents and transients) will show patterns of differentiation that reflect selection at functional loci. More broadly, we investigate the hypothesis that in addition to the process of genetic drift, disruptive selection is driving the differentiation of killer whale ecotypes in sympatry.

## Methodology

Samples were used from a long-term DNA archive built from previous studies (Hoelzel et al. 2007). Newly obtained samples from a population in the Southern Ocean at Marion Island were collected through remote biopsy sampling, using protocols approved by the University of Pretoria's Animal Use and Care Committee (EC023-10) and under permit from the Prince Edward Islands Management Committee. Details on sample numbers and origins are provided in Table S1. The distribution of sample sites is illustrated in Figure 1.

### *RAD Sequencing*

A modified RAD Seq protocol (Baird et al. 2008) was carried out using the NotI restriction enzyme to obtain genome-wide nuclear data (see Moura et al. 2014b for details). A subset of the samples (N=43) were used in an earlier study on phylogenomics (Moura et al. 2014b). Briefly, genomic DNA (500ng-1µg) was digested at 37°C overnight, followed by 5 minutes at 65°C for termination, with fragments isolated with AMPure XP beads (Agencourt). The forward adapter (p5 adapter) was modified to employ 4 adapters for initial ligation, thus facilitating sequencing cluster differentiation (see Moura et al. 2014b).

Adapter ligated fragments were sheared by sonication targeting a 500 bp average fragment size, recovered with AMPure XP beads and end repaired using a commercial kit (NEB E6050L) following manufacturer's instructions. Fragments were then selected using streptavidin magnetic beads (Dynabeads® M-280 Streptavidin cat no11205D Life Technologies), using 15 minutes of binding time at room temperature with agitation, and

washed 3 times with Tris/ Ethylenediaminetetraacetic acid (TE). DNA that bound to the beads was ligated to a universal p7 sequence adapter.

Amplification primers were designed with 8bp barcodes to allow multiplexing in a single sequencing lane. A single barcoded primer and a universal primer were used to amplify each sample. Cycling conditions included an initial denaturation step at 98°C for 30 seconds, followed by 12-14 cycles of one denaturation step at 98°C for 10 seconds, annealing at 60°C for 30 seconds, extension at 72°C for 30 seconds, followed by a final 5 minute extension step. Samples were purified with AMPure XP beads which were washed with 80% ethanol and resuspended in 10mM Tris pH 7.5. A Qubit (Life Technologies) was used for quantity assessment, and a Bioanalyser 2100 (Agilent) for quality assessment. A 1.5% Pippin prep cassette (Sage Scientific) was used to select pooled samples based on size and quantified by qPCR (KAPA). Two initial 2x150 MiSeq runs were carried out without phiX, for a pool of 5 libraries (one sample per library) using this modified and the original Baird et al. (2008) published approach, for quality control (see supplement to Moura et al. 2014b for more detail). The two methods provide comparable results. Sequencing of the pooled libraries was performed on the Illumina HiSeq 2000 using v3 chemistry.

#### *SNP Mapping and Genotyping*

Trimmed short reads were mapped against the killer whale genome version 1.1 (GenBank accession number ANOL000000000) using BWA short read mapper (Li and Durbin 2009). SNP genotypes were called with the Genome Analysis Toolkit (GATK) software



Accepted Article  
package (McKenna et al. 2010), using the Unified Genotyper module (DePristo et al. 2011).

A multi-sample genotype calling was carried out, including only variable SNPs that were called with a minimum base quality score of 10. VCFtools (Danecek et al. 2011) was then used to remove all indels, as well as positions with average coverage below 20X and genotype quality below Q20. SNPs for which at least one individual had missing data were removed from the datafile (so that we included only the SNPs that had been confidently scored for all individuals; input file provided on Dryad).

#### *Selection detection and population structure analyses*

The  $F_{ST}$  outlier method implemented in LOSITAN (Antao et al. 2008) was used to detect putative signs of selection. The following ecotypes/geographic locations were used to define populations (following the results obtained in Hoelzel et al. 2007; Parsons et al. 2013): Marion Island (MI), Iceland (IC), North Pacific Offshores (OS), Alaskan Residents (AR), Southern Residents (SR), Alaska Transients (AT), California Transients (CT), Bering Sea Residents (BS) and Russian Residents (RU). See table S1 for ecotype designation for each putative population. Baseline mean neutral  $F_{ST}$  was calculated excluding putative selected loci, which were detected using the bisection algorithm over 50,000 simulations with a false discovery rate of 0.1. Fixed differences were identified from genotype profiles and tested against chance using a binomial distribution test.

Comparisons of population structure based on SNPs were carried out using the DAPC (Discriminant Analysis of Principal Components) method (Jombart et al. 2010) implemented

Accepted Article

in the software package Adegenet (Jombart and Ahmed 2011) both by assigning each individual to the predefined populations, and by determining the best supported number of clusters (1-20) without *a priori* identification of structure. In both analyses, the first 40 principal components were retained, as were the first 4 discriminant functions. Patterns of population structure obtained with both neutral and positively selected SNPs were analyzed using factorial correspondence analysis (FCA) as implemented in Adegenet. Patterns of population diversity and differentiation between neutral and selected SNPs were investigated by comparing pairwise  $F_{ST}$  and nucleotide differences, using the software Arlequin (Excoffier et al. 2005). Differential patterns were assessed using the non-parametric Mann-Whitney U-test and  $\chi^2$  in 2x2 contingency tables (further details provided in the results section).

#### *Functional annotation of relevant SNPs*

In order to identify sites within the relatively well-annotated bottlenose dolphin genome, a sufficient region of sequence to permit confident alignment (2,000 bp) centred by the SNP of interest was retrieved from the killer whale genome and mapped against the bottlenose dolphin genome (version 1.69) using the Ensembl BLAT alignment tool. Functional annotation of the genes found to be physically linked to any of these SNPs was carried out using the DAVID web tool (Sherman et al. 2007). This analysis was done for all SNPs found to have fixed differences between populations, and for the SNPs with the strongest evidence for either positive or balancing selection based on the analysis using LOSITAN (limited to those assigned as an outlier with a probability of 1). Genes of interest were identified as those within 5 Kb of a SNP that was either a strong outlier or one of the

fixed differences. A minimum linkage block size of 5Kb has been suggested in some studies (e.g. Nash et al 2005; Shimada et al 2011). Due to potential inbreeding and small effective population sizes linkage blocks in killer whales may be relatively large, but the actual size is unknown. Therefore, for our study 5 Kb was chosen as a compromise: short enough for the SNP to be very likely within a linkage block, and long enough to include a reasonable number of loci to investigate.

To test whether any functions were over-represented for the genes linked to SNPs putatively under selection, we carried out the gene ontology (GO) term enrichment analysis as implemented in the software FatiGO + (Al-Sharour et al. 2007), using the Babelomics portal (Medina et al. 2010). This software compares the proportional representation of a given GO term in the submitted gene list to the proportion in a reference genome. Significance was assessed by means of a two-tailed Fisher exact test, and only those results still significant at the  $p < 0.05$  level after correction for type 1 errors were included.

Enrichment analyses were carried out against the human genome background list, with any duplicate entries removed. We chose the human background list because this is the mammalian genome with the most complete and robust level of annotation. However, we also assessed enrichment against a background list of genes linked to the SNPs identified from our RAD analysis. First we identified genes within 5 Kb of all SNPs identified in our RAD analysis, and from the list of genes associated with SNPs identified as neutral (identified using LOSITAN; see above), generated 10 random lists of the same length as the putative

selected loci list. These were investigated for the over-representation of GO terms using FatiGO+ using the human background list. On the assumption that the full RAD SNP-associated list would provide a representation of the background GO terms found across the full genome (though at a lower resolution), we also compared our putative selected loci list against GO terms for genes associated with the full set of RAD SNP loci.

### *Demographic analyses*

To investigate killer whale demographic history we analysed neutral SNPs using diffusion approximation of the joint allele frequency spectrum in *∂a∂i* version 1.6.3 (Gutenkunst et al. 2009). To polarize the SNPs and determine the derived allele, killer whale reads were mapped against the bottlenose dolphin (*Tursiops truncatus*) genome assembly version 1.69 available from the Ensembl database. SNPs were called and filtered as described above, and sites that were found to be under positive or balancing selection were removed. A statistical correction for errors in the outgroup *Tursiops* sequence was applied using a trinucleotide substitution matrix from Hwang and Green (2004) and an estimate of 1.82% sequence divergence between the dolphin and killer whale. Our estimate of sequence divergence was based on a genome size of 2.3 Gb, a divergence time of 10 million years (following the tmrca between the two species as determined from a multilocus phylogeny, McGowen et al. 2009), and a mutation rate of  $2.34 \times 10^{-8}$  substitutions/nucleotide/ generation, following the mutation rate estimates for the nuclear genome in Odontocetes (Dornburg et al. 2012).

The *∂a∂i* software can model a maximum of three populations at a time. Therefore we focused on two trios of populations that reflected the most informative lineage histories as indicated by the topology in our phylogenies from Moura et al. (2014b). The first analysis included the Southern Ocean Marion Island (MI), Alaska Transient (AT), and Offshore (OS) populations, and the second included AT, OS, and the Alaska Resident population (AR). We selected AT and AR for these analyses because they represented the largest samples for the transient and resident ecotypes, respectively, and are found in sympatry. In our model, the ancestral population was allowed to vary in population size until the initial population split, and then the two daughter populations were allowed to vary in size with symmetric migration, until the second divergence, after which all three populations were allowed to vary in size and experience symmetric migration. To investigate the order in which populations split, we shuffled the order of the populations for each analysis and then selected the model that demonstrated the maximum likelihood. We then performed 25 independent replicates for that model, chose the maximum likelihood outcome, and report the estimates for all of the parameters under that model (shown as 'estimate' in Tables S2 & S3). We assessed uncertainty by conducting 100 bootstrap replicates with replacement for that model, optimizing parameters for 25 iterations each. Bootstrap parameter distributions were mostly right skewed, and therefore we report median values together with parametric parameters for log(10) transformed data (mean and 95% confidence intervals). We discarded runs that failed to converge (after Excoffier et al. 2013), including all runs that failed to complete or where parameter values were continually increasing. Parameter estimates were converted to more meaningful demographic units by applying the mutation rate specified above, a sequence length of 1.45 Mb, and a generation time of 25.7 years (Taylor et al. 2007).

## Results

### *Neutral Population Structure*

A total of 3281 variable SNPs could be confidently mapped to the killer whale reference genome (see methods), shared among 115 individuals (Genbank accessions: SAMN03020306-SAMN03020378; SAMN02820869-SAMN02820892; SAMN02820894-SAMN02820911). Using LOSITAN (Antao et al. 2008), 347 SNPs were identified as outliers, 168 as being potentially under positive selection and 179 as being under balancing selection (Figure 2). Discriminate Analysis of Principal Components (DAPC, Jombart et al. 2010) using only the neutral SNPs was able to correctly discriminate all ecotypes and geographical populations when population identity was included as a factor (Figure 3a). When there was no *a priori* population identification only the Russian and Bering Sea resident populations grouped together (Figure 3b). From both  $F_{ST}$  (Table 1) and FCA (Figure 4a) a hierarchical structure becomes apparent, such that ecotype clusters were more strongly identified than the differences among populations within ecotypes (Table 1, Figure 3a). However, FCA could still discriminate population structure within ecotypes when they were analysed on their own (Figure S1). This pattern is reinforced by comparisons based on pairwise allelic differences (Figure 5). Figure 5 provides a heat map representation of this metric for all pairwise population comparisons, and shows the distinction between comparisons among populations based on neutral loci (above diagonal) and those potentially under selection (below diagonal).

## *Demographic inference*

We assessed demographic history using neutral loci to better understand the process of differentiation among these populations. Our model testing in *∂a∂i* (Gutenkunst et al. 2009) compared two trios of populations, Marion Island (MI), Alaskan transients (AT) and Offshores (OS); and AT, OS and Alaskan residents (AR; see methods). MI was the ancestral population associated with the maximum likelihood for the first trio compared, and AT for the second (Tables S4 & S5; consistent with Moura et al. 2014a; 2014b). Demographic analyses suggest that the common ancestor of MI, AT and OS was initially small and then grew before the MI population split. Both lineages then continued to grow until later declining, with the AT/OS lineages also declining (Figure 6a; see Table S2 for  $N_e$  values and confidence limits). Migration during the two population period was relatively high and then declined after the split into three populations (Table S2).

In the second trio (Figure 6b; Table S3), the ancestral population grew until the split between AT and OS/AR, then AT declined (similar to the result from the first trio comparison). The OS/AR lineage grew until the split when both OS and AR declined. Migration was again higher during the 2-population phase, declining after that. The timeframe of all modelled population divisions is estimated to be within the late Pleistocene or the Holocene (Tables S2 & S3), with the most recent period reflecting population decline and reduced migration. Details of all parameter estimates are given in Tables S2 & S3, allele frequency spectra for maximum likelihood (ML) runs are shown in Figures S2 and S3, and the model fit is shown in illustrations in Figures S2 and S3, with the residuals (rows 3 and 4) indicating a good fit.

SNPs identified as being under positive selection (showing high  $F_{ST}$  for a given level of diversity) generated distinct patterns of population structure compared to neutral markers. The FCA analyses of neutral loci suggest shared kinship between offshores, the Icelandic population and transients, as suggested in earlier studies (Pilot et al. 2010), with the Marion Island population being most differentiated. For the positive outlier loci offshores and Iceland become more differentiated and Marion Island and transients more similar (Figure 4). We tested for quantitative evidence of distinct differentiation patterns comparing neutral and outlier loci, with respect to the two main ecotypes in the North Pacific, residents and transients. For measures of  $F_{ST}$  we consider the following ratio for all possible pairwise combinations: the value of  $F_{ST}$  for population comparisons between resident and transient ecotypes/ the value of  $F_{ST}$  for population comparisons within resident or transient ecotypes. For neutral loci this average ratio is 4.15 (range = 1.34 – 8.59,  $N = 24$ ), significantly lower than when only outlier loci are included (8.18, range = 2.61 – 14.48,  $N = 24$ ,  $Z = -2.96$ ,  $p = 0.003$ ; Mann-Whitney U-test).

This same pattern was evident based on pairwise nucleotide differences (as illustrated in Figure 5) such that the magnitude of the ratio between ecotypes compared to within ecotypes was greater for the outlier loci (3.6 fold compared to 1.6 fold;  $\chi^2 = 51.06$ ,  $p < 0.0001$ , d.f. = 1; Yate's correction applied). Comparing residents and transients against the other three groups combined (OS, IC and MI) for pairwise nucleotide differences also showed a distinct pattern for neutral compared to outlier loci. There was greater similarity



between residents and the outgroups for neutral markers, but greater similarity between transients and the outgroups for outlier loci (Figure 5;  $\chi^2 = 75.4$ ,  $p < 0.0001$ , d.f. = 1).

Fixed differences (marked with an 'X' in Figure 2) were found for both neutral and positive outlier loci in pairwise comparisons between putative populations, but were much more common for outlier loci (Table 2). This will likely include outliers due to both selection and drift (see discussion). Overall there were 33 SNPs showing fixed differences for one or more population comparisons, however most sample sizes were small, and so only some showed this more often than may occur by chance (Tables 2 & S6). From FCA analysis of these SNPs, a pattern of four population clusters emerged: the North Pacific resident populations, the North Pacific transients together with the Marion Island whales, the offshores, and the Icelandic sample (Figure S4). There were more fixed difference loci than expected by chance (Table 2) for comparisons between populations from each of these clusters, however the number between residents and transients was relatively low and poorly supported. Among resident populations SR (the 'southern residents' from Washington State waters) was most differentiated (Table 2), and also showed the lowest level of within population differentiation for both neutral and outlier loci (Figure 5).

#### *Functional annotation of outlier and fixed difference SNPs*

We considered possible linkage to known genes for the most extreme outliers (red dots in Figure 2; assigned as an outlier with a probability of 1; see methods) for loci under putative positive or balancing selection. Out of 21 SNPs identified for balancing selection,

only 3 loci were closely linked (defined here as within 5Kb, see methods) to a known gene, and these were associated especially with cell signalling (RAB43, BHLHB9, GRM7). Among those identified for positive selection a higher proportion (11/19) were linked to genes associated with functions relevant to reproduction, growth and metabolism, among others (see Table S7). The genes linked to fixed difference SNPs again represented loci relevant to aspects of physiology, growth and reproduction (Table S8). There was just one locus where the fixed difference represented a non-synonymous change within an exon. This was for GATA4 (at residue 187 changing a proline to a glutamine), a transcription factor involved in various functions including the development of cardiac tissue (Hu et al. 2010) and testes size/reproductive function (Kyrölähti et al. 2011). The fixed difference at this locus was between the resident populations and all others.

We then considered the gene ontology of these loci (all 24 unique loci identified as under strong positive selection or reflecting fixed differences as identified above) in the context of the proportional representation of gene ontology (GO) terms compared to what may be expected by chance from a fully annotated reference genome (using the human genome due to the high quality annotation; Table S9). The results show that 30 terms are significantly enriched compared to what would be expected by chance, including terms associated with the digestive and circulatory systems. Significantly enriched functions and the associated genes are listed in Table S9 and the relevant genes marked in bold text in Tables S7 and S8. Two transcription factor genes were associated with multiple enriched functional terms, GATA4 and MZF1.

Our tests for over-representation of GO terms in 10 random lists of genes linked to neutral loci (each of the same length as the list of genes putatively under selection) found no evidence of over-representation for a p-value after correction for false discovery of either 0.05 or 0.10. When we used genes associated with all SNP loci identified from our RAD analysis as the background instead of the full human genome, only 4 terms were over-represented at a p-value of 0.05, but 32 at a p-value of 0.10, indicating that this comparison provided reduced power. GATA4 and MZF1 remained associated with multiple enriched terms.

## Discussion

### *Population structure and demographic analyses*

Summary statistics ( $F_{ST}$  and pairwise differences) and ordination analyses (DAPC and FCA) show strong differentiation both among ecotypes and among geographic populations within ecotypes at neutral loci, providing a finer level of resolution to the patterns found earlier in microsatellite DNA studies (Hoelzel et al. 2007; Parsons et al. 2013). In particular, in contrast to the earlier microsatellite DNA studies, all putative populations, including four parapatric resident populations within the North Pacific, can be genetically distinguished.

There was also evidence for hierarchical clusters that had not been clearly defined using microsatellite DNA data, and that sometimes differ from inference previously gained from mtDNA genetic lineages (Hoelzel et al. 2002; Morin et al. 2010; Moura et al. 2014b). This inconsistency between inference from mtDNA and nuclear markers (e.g. a level of

Accepted Article

mtDNA similarity between residents and offshores not seen for nuclear markers) had been noted in earlier studies using alternative methods (Pilot et al. 2010, Moura et al. 2014b), and is likely due to stochastic factors affecting the mtDNA lineages. The most coherent ordination clusters grouped populations by resident or transient ecotype in the North Pacific. Offshores, Iceland and Marion Island were all clearly differentiated from the resident and transient clusters based on neutral loci. Demographic analyses concurred with the ordination analyses, as well as previous studies, indicating low migration rates between the sympatric transient and resident ecotypes.

Our temporal estimates of population differentiation (Tables S2 & S3) are broadly consistent with earlier microsatellite DNA estimates suggesting that populations in the North Pacific differentiated during the relatively recent past (during the late Pleistocene and Holocene, Hoelzel et al. 2007). They support previous studies suggesting a higher mutation rate (3%/ MY for the mitogenome; Ho & Lanfear 2010; Moura et al 2013b) than used in published estimates of the time of initial division between transients in the North Pacific and other lineages worldwide (702 Ka based on 0.3%/ MY; Morin et al. 2010; Foote et al. 2011). Our calculation of divergence times is based on a published estimate for the average substitution rate in the Odontocete nuclear genome (see Methods), and is more consistent with inference using the faster mtDNA mutation rate estimate. The timing is consistent with a signal for a population bottleneck during the last glacial period, as indicated from whole genome analyses for individual killer whales from the North Pacific and North Atlantic (Moura et al. 2014a).

The demographic data presented in Figure 6 suggest a smaller ancestral population (possibly post-bottleneck; Moura et al. 2014a) that was expanding at the time of the division between the Southern Ocean and North Pacific transient populations. If new habitat was being released (as would be the case following a glacial period) at the time of the establishment of the nearshore (resident) piscivorous ecotype, then this would be consistent with the evident later reduction in effective population size as founder populations were established. Low gene flow, kin associations in social groups, and the expansion of these new populations along matrilineal lines would reinforce and exaggerate the degree of differentiation between the source and the newly established populations. The relatively high estimated level of gene flow after the initial split between MI and AT, and the subsequent reduction in gene flow, are consistent with this scenario and also consistent with recent divergence within the North Pacific. Inference from allele frequency spectra (as implemented in *daði*) has known limitations (Myers et al. 2008), but also empirical support in comparison with known geologic or historical events (Gutenkunst et al. 2009; Zhao et al. 2013). In our case comparable values are similar to those obtained by different methods in earlier studies (e.g. Hoelzel et al. 2007; Moura et al. 2014a).

#### *Differentiation at putative selected SNPs*

The pattern of differentiation among putative populations changes when positive outlier loci are considered on their own. For example, there is increased similarity among separate populations within each of the resident and transient ecotypes, and significantly greater differentiation between these two ecotypes (see Table 1 & Figure 5). A similar pattern was seen for genotypes associated with freshwater or marine habitat in

Accepted Article

sticklebacks, such that loci under selection were more similar within ecotypes (Hohenlohe et al. 2010). However, the method chosen to identify outlier loci (LOSITAN) has a high false discovery rate (Antao et al. 2008), which means that there are likely neutral loci among those identified as being under selection (while the identification of neutral loci should be relatively free of loci under selection). Therefore, the observed pattern could be expected either when loci under selection reflect genes that are associated with adaptation by ecotype, or if strong drift generates differentiation between ecotypes more than among other populations, or by some combination of both factors.

Because ecotype populations are among those used in identifying the outliers, an exaggerated pattern of drift could be seen in the outliers, though this effect should be similar for neutral loci. However, the comparison of each North Pacific ecotype (residents and transients) against outgroups (Marion Island, Iceland and the Offshores) provided additional inference that there was exaggerated differentiation for outliers which more likely reflects differential selection among ecotypes. In this case the outliers showed the opposite relationship to the outgroups compared to the neutral markers (Figure 5). It is also the case that strong outliers and fixed differences were associated with genes that have ecologically relevant functions (Tables S7-S9). Because of the potential for neutral loci being grouped with the outliers, we considered gene ontology terms only for those outliers that were strongly supported (as identified by the program LOSITAN, see methods), or associated with fixed differences between populations.

Killer whale ecotypes forage on different resources (though some populations may prey on multiple resource types; see Best et al. 2010, Foote et al. 2013), may have different metabolic requirements as a consequence, and have different social structure (e.g. Ford et al. 1998; Beck et al. 2012), which may have implications for reproductive strategy including the potential for sperm competition. A total of 24 loci potentially under positive selection linked to SNPs were identified, and some of these did have functions consistent with relevant aspects of digestive physiology and reproductive biology (see Tables S7 & S8). Because this species has a demographic history involving population bottlenecks (see Hoelzel et al. 2002, Moura et al. 2014a), and a social structure that promotes kin association (though mating is typically outside of the social group, and kinship within social groups not substantially different than among regional social groups; Pilot et al. 2010, Ford et al. 2011), linkage blocks may be long. However, since an assessment of enrichment showed that some GO terms associated with genes from our list were significantly over-represented (Table S9), the implication is that we are identifying loci that are relevant to selection. This is reinforced by the fact that random lists of genes linked to neutral SNPs showed no evidence for enrichment of GO terms. For the list of genes putatively under selection there was significantly stronger over-representation than expected by chance for 30 GO terms, including terms associated with digestion (consistent with one key distinction among ecotypes).

We used the human genome as a reference as it is the best annotated mammalian genome and GO terms should be mostly species neutral (see Primmer et al. 2013). At the same time, the choice of an appropriate background reference list is important (Primmer et

al. 2013), and in this case we contend that the full set of mammalian genes is appropriate, even though the outliers and fixed difference SNPs were chosen from among the loci identified from a subsample of sites provided by the RAD data. This subset provides a list that is expected to be representative, but is an order of magnitude smaller, and therefore only an approximate representation. When we compared against the RAD list we found that there is lower power as expected, and although some GO terms associated with the same subset of genes are identified as potentially over-represented, the resolution is too low for strong inference. When using the full human reference there is a risk that enrichment would represent false positives, but that would imply finding enriched terms for our sets of random neutral loci, which we did not observe.

In general, differential patterns for positive outliers compared to neutral loci imply selection (for at least some loci among the outliers), though it was not possible to fully determine the nature of the adaptive differences. Distinct structure at outlier loci included Marion Island being more similar to the transient populations, and Iceland (and to some extent offshores) being more distinct from the residents. These are in each case comparisons within broad ecotype categories (since Marion Island whales take pinnipeds as part of their diet as do transients, while offshores, Icelandic and resident whales all feed on fish, see review in de Bruyn et al. 2013). While the implication from the increased similarity between transients and Marion Island is consistent with shared ecotype, increased differentiation among the three fish-eating groups implies that there are additional factors beyond prey choice involved in the adaptation of these various populations to their different habitats (though phenotypic variation among these populations is not strong, and



we don't know enough about the relevant genes to speculate further). The possibility of changing specialisations over time in the North Atlantic would potentially be relevant (Foote et al. 2013). More generally, it is clear that it will be important to learn more about the characteristics that may reflect ecotype adaptation. Increasingly studies have found that key differences in phenotype are likely to be related to gene regulation (Carroll 2005), and some of our results (e.g. with respect to GATA4 and MZF1) could be consistent with that.

#### *Fixed differences and the question of taxonomy*

Some further insight can be gained through the analysis of fixed differences among populations. These are differences that could be sustained either through an extended period of isolation and drift, by inbreeding, by strong selection, or by sampling effects from small sample sizes. From our data there is much stronger evidence for fixed differences at outlier than neutral loci, but these will likely include loci that are outliers due to drift as well as selection. The pattern of differentiation at the 33 fixed loci (based on FCA) emphasizes the clusters defined by the outlier loci in general. Transients cluster with Marion Island, residents group together, and both offshores and Iceland form independent clusters (Figure S4, Table 2). Among the fixed differences, 20 were within 5 Kb of a functional locus with 16 being within or just outside the gene. The gene ontology terms indicate functions associated with digestion, nutrient metabolism, protein metabolism, body development, and reproduction (among other potentially relevant functions; see Table S8).

Accepted Article

Just one locus showed a fixed difference at a non-synonymous change within an exon, the transcription factor GATA4, and this reflected a fixed difference between the residents and all other groups. Among its various functions, GATA4 is involved in testis development in mice (Kyrönlahti et al. 2011). The highly stable social groups and apparent strategy for mating in temporary assemblages (Pilot et al. 2010) within the resident populations may be consistent with sperm competition, possibly promoted by changes at this locus. However, there are few relevant data on sperm competition in this species, and among 31 cetacean species the killer whale was ranked 13<sup>th</sup> for residual testes size (used as an index for the potential for sperm competition; MacLeod 2010). More work is needed to further explore the potential role of sperm competition.

The presence of fixed differences can raise questions about alpha taxonomy, although there is at this stage no set expectation for the proportion of fixed differences that may define a species-level difference, and the range of observed values is broad. Empirical observations for allozyme markers suggested that distinct species can be expected to exhibit fixed differences at approximately 15% of loci analysed (O'Donnell et al. 2004). However, a genomic study comparing two closely related flycatcher species of the genus *Ficedula* (with a 2Ma most common recent ancestor, but still exhibiting limited hybridization) found that only 1.8% of all segregating sites identified were fixed between the two species (Ellegren et al. 2012). The percentages of fixed differences between killer whale populations were typically much smaller (e.g. average values for resident populations vs Iceland = 0.328%; residents vs all others=0.177%; residents vs transients = 0.103%). There is only one SNP showing a fixed difference between all individuals of one ecotype (residents) and all others,

and that was GATA4 (which also shows a non-synonymous change). The southern residents show more pronounced fixed differences (Table 2) and less differentiation within the population compared to other putative populations (Pilot et al. 2010; Figure 5). Earlier studies using non-equilibrium models showed support for gene flow between residents and transients following a point of division (Hoelzel et al. 2007), however our RAD data suggest that this may have been mostly during the period soon after division (consistent with the temporal data about migration shown in Table S3).

### *Conclusions*

As discussed in earlier studies (Hoelzel et al. 2007), regional killer whale populations can be small and evidently composed of an extended matriline, likely founded by one or a few stable social groups. The southern residents are an extreme example of this. For this reason, it is appropriate to be cautious about the interpretation of fixed differences, as these could be promoted relatively quickly by drift when  $N_e$  is small. Earlier studies have proposed everything from relatively recent (Holocene) population differentiation (e.g. Hoelzel et al. 2007) to species-level differences established in the middle Pleistocene (Morin et al. 2010), and we do not propose to fully resolve the question of alpha taxonomy here. However, the timeframe implied from our RAD data and from the demographic profiles generated from nuclear genomes (Moura et al. 2014a) suggest relatively recent events. The data on population structure based on outlier loci and the significant over-representation of biologically relevant gene ontology terms suggest that natural selection has contributed to the degree of differentiation among populations of different ecotypes in the North Pacific. Taken together these data suggest that differentiation in sympatry is based in part on

ecological processes, but that differentiation is likely being facilitated by the life history of killer whales, founder events and differentiation by drift. Both the demographic analyses and the small proportion of fixed differences suggest that if incipient speciation is the correct interpretation, it began very recently and is ongoing.

### **Acknowledgements:**

This study was funded by the Natural Environment Research Council UK (grant number NE/014443/1). The Department of Environmental Affairs supplied logistic support within the South African National Antarctic Programme for work done at Marion Island. Financial support for field assistance was provided by the South African Department of Science and Technology, through the National Research Foundation (NRF), in support of the Marine Mammal Programme of the MRI. We thank the Baylor College of Medicine and the Centre for GeoGenetics in Copenhagen for provision through NCBI of the ANOL000000000 genome sequence for reference prior to its publication. We thank Ken Balcomb, Mike Bigg, Nancy

Black, Vladimir Burkanov, Andy Dizon, Dave Ellifrit, Mike Ford, Jeff Foster, Rich Osborne, Jen Schorr, and Jóhann Sigurjónsson for earlier contributions to our killer whale DNA archive.

### **References**

Alvarado Bremer JR, Viñas J, Mejuto J, Ely B, Pla C. 2005. Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression,

and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Mol Phylogenet Evol.* 36: 169-187.

- Al-Shahrour F, Minguez P, Tarraga J, Medina I, Alloza E, Montaner D, Dopazo J. 2007. FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res.* 35: W91-W96
- Amaral AR, Beheregaray LB, Bilgmann K, Boutov D, Freitas L, Robertson KM, Sequeira M, Stockin KA, Coelho MM, Moller LM. 2012. Seascape genetics of a globally distributed, highly mobile marine mammal: the short-beaked common dolphin (genus *Delphinus*). *PLoS One.* 7: e31482.
- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. 2008. LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics.* 9: 323.
- Atarhouch T, Rami M, Naciri M, Dakkak A. 2007. Genetic population structure of sardine (*Sardina pilchardus*) off Morocco detected with intron polymorphism (EPIC-PCR). *Mar Biol.* V150: 521-528.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One.* 3: e3376.
- Beck S, Kuningas S, Esteban R, Foote AD. 2012. The influence of ecology on sociality in the killer whale. *Behav. Ecol.* 23: 246-253.
- Beheregaray LB, Sunnucks P. 2001. Fine-scale genetic structure, estuarine colonization and incipient speciation in the marine silverside fish *Odontesthes argentinensis*. *Mol Ecol.* 10: 2849-2866.
- Bierne N, Bonhomme F, David P. 2003. Habitat preference and the marine-speciation paradox. *P Roy Soc London B Biol Sci.* 270: 1399-1406.
- Bowen BW, Rocha LA, Toonen RJ, Karl SA. 2013. The origins of tropical marine biodiversity. *Trends Ecol Evol.* 28: 359-366.
- Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biology.* 3: e245.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics.* 27: 2156-2158.
- de Bruyn M, Nugroho E, Hossain MM, Wilson JC, Mather PB. 2005. Phylogeographic evidence for the existence of an ancient biogeographic barrier: the Isthmus of Kra Seaway. *Heredity.* 94: 370-378.
- de Bruyn PJN, Tosh CA, Terauds A. 2013. Killer whale ecotypes: is there a global model? *Biol Rev.* 88: 62-80.
- Deecke VB, Ford JKB, Slater PJB. 2005. The vocal behaviour of mammal-eating killer whales: communicating with costly calls. *Anim Behav.* 69: 395-405.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 43: 491-498.
- Domingues VS, Santos RS, Brito A, Alexandrou M, Almada VC. 2007. Mitochondrial and nuclear markers reveal isolation by distance and effects of Pleistocene glaciations in the northeastern

Atlantic and Mediterranean populations of the white seabream (*Diplodus sargus*, L.). *J Exp Mar Bio Ecol.* 346: 102-113.

- Domingues VS, Stefanni S, Brito A, Santos RS, Almada VC. 2008. Phylogeography and demography of the Blenniid *Parablennius parvicornis* and its sister species *P. sanguinolentus* from the northeastern Atlantic Ocean and the western Mediterranean Sea. *Mol Phylogenet Evol.* 46: 397-402.
- Dornburg A, Brandley MC, McGowen MR, Near TJ. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Mol Biol Evol.* 29: 721-736.
- Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, Uebbing S, Wolf JBW. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature.* 491: 756-760.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online.* 1: 47-50.
- Foote AD, Morin PA, Durban JW, Willerslev E, Orlando L, Gilbert MTP. 2011. Out of the Pacific and back again: insights into the matrilineal history of Pacific killer whale ecotypes. *PLoS One.* 6: e24980.
- Foote AD, Newton J, Avila-Arcos MC, Kampmann M-L, Samaniego JA, Post K, Rosing-Asvid A, Sinding MS, Gilbert MTP. 2013. Tracking niche variation over millennial timescales in sympatric killer whale lineages. *Proc. Royal Soc.* 280: 20131481.
- Ford JKB, Ellis GM, Balcomb KC. 2000. Killer Whales: The Natural History and Genealogy of *Orcinus orca* in British Columbia and Washington State. Vancouver: UBC Press.
- Ford JKB, Ellis GM, Barrett-Lennard LG, Morton AB, Palm RS, Balcomb KC. 1998. Dietary specialization in two sympatric populations of killer whales (*Orcinus orca*) in coastal British Columbia and adjacent waters. *Can J Zool.* 76: 1456-1471.
- Ford MJ, Hanson MB, Hempelmann JA, Ayres KL, Emmons CK, Schorr GS, Baird RW, Balcomb KC, Wasser SK, Parsons KM, et al. 2011. Inferred paternity and male reproductive success in a killer whale (*Orcinus orca*) population. *J. Hered.* 102: 537-553.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5: e1000695.
- Ho SY, Lanfear R. 2010. Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondr. DNA* 21:138-146.
- Hoelzel AR. 2009. Evolution of population structure in marine mammals. In: Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C, editors. Population Genetics for Animal Conservation. Cambridge: Cambridge University Press.
- Hoelzel AR, Hey J, Dahlheim ME, Nicholson C, Burkanov V, Black N. 2007. Evolution of population structure in a highly social top predator, the Killer Whale. *Mol Biol Evol.* 24: 1407-1415.

- Hoelzel AR, Natoli A, Dahlheim ME, Olavarria C, Baird RW, Black NA. 2002. Low worldwide genetic diversity in the killer whale (*Orcinus orca*): implications for demographic history. *P Roy Soc B Biol Sci.* 269: 1467–1473.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6: e1000862.
- Hu D-L, Chen F-K, Liu Y-Q, Sheng Y-H, Yang R, Kong X-Q, Cao K-J, Gu H-T, Qian L-M. 2010. GATA-4 promotes the differentiation of P19 cells into cardiac myocytes. *Int J Mol Med.* 26: 365-372.
- Hwang DG, Green P. 2004. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc Natl Acad Sci U S A.* 101: 13994-14001.
- Jombart T, Ahmed I. 2011. Adegnet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics.* 27: 3070-3071.
- Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11: 94.
- Jorgensen HBH, Hansen MM, Bekkevold D, Ruzzante DE, Loeschcke V. 2005. Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. *Mol Ecol* 14: 3219-3234.
- Krahn MM, Herman DP, Matkin CO, Durban JW, Barrett-Lennard L, Burrows DG, Dahlheim ME, Black N, LeDuc RG, Wade PR. 2007. Use of chemical tracers in assessing the diet and foraging regions of eastern North Pacific killer whales. *Mar Environ Res.* 63: 91-114.
- Kyrönlähti A, Euler R, Bielinska M, Schoeller EL, Moley KH, Toppari J, Heikinheimo M, Wilson DB. 2011. GATA4 regulates Sertoli cell function and fertility in adult male mice. *Mol Cell Endocrinol.* 333: 85-95.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754-1760.
- MacLeod CD. 2010. The relationship between body mass and relative investment in testes mass in cetaceans: implications for inferring interspecific variations in the extent of sperm competition. *Mar Mamm Sci.* 26:370–380.
- McGowen MR, Spaulding M, Gatesy J. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol.* 53: 891-906.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20: 1297-1303.
- Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, Conesa A, Tarraga J, Pascual-Montano A, Nogales-Cadenas R, Santoyo J, Garcia F, Marba M, Montaner D, Dopazo J. 2010. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res.* 38: W210-W213

- Mendez M, Subramaniam A, Collins T, Minton G, Baldwin R, Berggren P, Särnblad A, Amir OA, Peddemors VM, Karczmarski L, Guissamulo A, Rosenbaum HC. 2011. Molecular ecology meets remote sensing: environmental drivers to population structure of humpback dolphins in the Western Indian Ocean. *Heredity*. 107: 349-361.
- Möller L, Valdez FP, Allen S, Bilgmann K, Corrigan S, Beheregaray LB. 2010. Fine-scale genetic structure in short-beaked common dolphins (*Delphinus delphis*) along the East Australian Current. *Mar Biol*. 158: 113-126.
- Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Abel Nielsen SC, Rasmussen M, Willerslev E, Gilbert MT, Harkins T. 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res*. 20: 908-916.
- Moura AE, Natoli A, Rogan E, Hoelzel AR. 2013a. Atypical panmixia in a European dolphin species (*Delphinus delphis*): implications for the evolution of diversity across oceanic boundaries. *J Evol Biol*. 26: 63-75.
- Moura AE, Nielsen SCA, Vilstrup JT, Moreno-Mayar JV, Gilbert MTP, Gray H, Natoli A, Möller L, Hoelzel AR. 2013b. Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change. *Syst. Biol*. 62: 865-877
- Moura AE, Janse van Rensburg C, Pilot M, Tehrani A, Best PB, Thornton M, Plon S, de Bruyn PJN, Worley KC, Gibbs RA, Dahlheim ME, Hoelzel AR 2014a. Killer Whale Nuclear Genome and mtDNA Reveals Widespread Population Bottleneck During the Last Glacial Maximum. *Mol. Biol. Evol.* 31: 1121-1131
- Moura, A.E., Kenny, J.G., Chaudhuri, R., Hughes, M.A., Reisinger, R.R., de Bruyn, P.J.N., Dahlheim, M.E., Hall, N., Hoelzel, A.R. 2014b. Phylogenomics of the killer whale indicates ecotype divergence in sympatry. *Heredity* doi:10.1038/hdy.2014.67
- Myers S, Fefferman C, Patterson N. 2008. Can one learn history from the allelic spectrum? *Theor Popul Biol*. 73: 342-348.
- Naciri M, Lemaire C, Borsa P, Bonhomme F. 1999. Genetic study of the Atlantic/Mediterranean transition in sea bass (*Dicentrarchus labrax*). *J Hered*. 90: 591-596.
- Nash D, Nair S, Mayxay M, Newton PN, Guthmann J-P, Nosten F, Anderson TJC. 2005. Selection strength and hitchhiking around two anti-malarial resistance genes. *Proc. Royal Soc. B*. 272: 1153-1161.
- Natoli A, Birkun A, Aguilar A, Lopez A, Hoelzel AR. 2005. Habitat structure and the dispersal of male and female bottlenose dolphins (*Tursiops truncatus*). *P Roy Soc B Biol Sci.* 272: 1217-1226.
- Natoli A, Cañadas A, Peddemors VM, Aguilar A, Vaquero C, Fernández-Piqueras P, Hoelzel AR. 2006. Phylogeography and alpha taxonomy of the common dolphin (*Delphinus* sp.). *J Evol Biol*. 19: 943-954.
- O'Donnell K, Ward TJ, Geiser DM, Corby Kistler H, Aoki T. 2004. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. *Fungal Genet Biol*. 41: 600-623.



- Palumbi SR, Lessios HA. 2005. Evolutionary animation: how do molecular phylogenies compare to Mayr's reconstruction of speciation patterns in the sea? *Proc Natl Acad Sci U S A*. 102: 6566-6572.
- Parsons KM, Durban JW, Burdin AM, Burkanov VN, Pitman RL, Barlow J, Barrett-Lennard LG, LeDuc RG, Robertson KM, Matkin CO, Wade PR. 2013. Geographic patterns of genetic differentiation among killer whales in the northern North Pacific. *J Hered*. 104: 737-754.
- Pilot M, Dahlheim ME, Hoelzel AR. 2010. Social cohesion among kin, gene flow without dispersal and the evolution of population genetic structure in the killer whale (*Orcinus orca*). *J Evol Biol*. 23: 20-31.
- Primmer CR, Papakostas S, Leder EH, Davis MJ, Ragan MA. 2013. Annotated genes and nonannotated genomes: cross-species use of gene ontology in ecology and evolution research. *Mol. Ecol*. 22: 3216-3241.
- Puckridge M, Last PR, White WT, Andreakis N. 2012. Phylogeography of the Indo-West Pacific maskrays (Dasyatidae, Neotrygon): a complex example of chondrichthyan radiation in the Cenozoic. *Ecol Evol*. 3: 217-232.
- Rousset F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour*. 8: 103-106.
- Ruzzante DE, Taggart CT, Cook C, Goddard S. 1996. Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: microsatellite DNA variation and antifreeze level. *Can J Fish Aquat Sci*. 53: 634-645.
- Sherman BT, Huang D, Tan Q, Guo Y, Bour S, Liu D, Stephens R, Baseler MW, Lane HC, Lempicki RA. 2007. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics*. 8: 426.
- Shimada Y, Shikano T, Merila J. 2011. A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Mol. Biol. Evol*. 28: 181-193.
- Taylor BL, Chivers SJ, Larese JP, Perrin WF. 2007. Generation length and percent mature estimates for IUCN assessments of cetaceans. *NOAA, NMFS, Southwest Fisheries Science Center Administrative Report LJ-07-01*: 21.
- Tringali M. 1999. Molecular Phylogenetics and Ecological diversification of the transisthmian fish genus *Centropomus* (Perciformes: Centropomidae). *Mol Phylogenet Evol*. 13: 193-207.
- White TA, Stamford J, Rus Hoelzel A. 2010. Local selection and population structure in a deep-sea fish, the roundnose grenadier (*Coryphaenoides rupestris*). *Mol Ecol*. 19: 216-226.
- Yoshida H, Yoshioka M, Shirakihara M, Chow S. 2001. Population structure of finless porpoises (*Neophocaena phocaenoides*) in coastal waters of Japan based on mitochondrial DNA sequences. *J Mammal*. 82: 123-130.
- Yurk H, Barrett-Lennard L, Ford JK, Matkin CO. 2002. Cultural transmission within maternal lineages: vocal clans in resident killer whales in southern Alaska. *Anim Behav*. 63: 1103-1119.
- Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, Hu Y, He W, Zhang S, Fan W, Zhu L, Li D, Zhang X, Chen Q, Zhang H, Zhang Z, Jin X, Zhang J, Yang H, Wang J, Wang J, Wei F. 2013. Whole-genome

sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet.* 45: 67-71.

**Data Accessibility:**

DNA sequences: Genbank accessions: SAMN03020306-SAMN03020378; SAMN02820869-SAMN02820892; SAMN02820894-SAMN02820911

VCF files for SNP data: Dryad doi: 10.5061/dryad.qk22t

**Figure Legends**

Figure 1 – Map of sample sites (colour coded to match Figures 4&7) and sample sizes parenthetically. Location abbreviations are as defined in Table 2.

Figure 2 - LOSITAN plot identifying outlier SNPs based on neutral expectations. X's represent loci with fixed differences between two or more populations. Red dots indicate the best supported outlier loci (used for the gene ontology analyses). Het stands for heterozygosity.

Figure 3 - DAPC for neutral loci with a) population identity assigned and b) when the program assigns clusters (see text for detailed parameters). Each bar represents a different individual, and red indicates stronger assignment. Location abbreviations are as defined in Table 2.

Figure 4 - FCA analysis of a) neutral loci and b) positive outliers.

Figure 5 – Heat map comparison of the number of allelic differences between neutral (above diagonal) and positively selected SNPs (below diagonal) within (along diagonal) and between populations identified in Figure 1. The colour bars indicate the range of values for neutral (above the scale bars) and outlier (below) SNPs.

Figure 6 – Analysis of demographic history of killer whales. a) Analysis including Southern Ocean Marion Island (MI) population, Alaska Transients (AT), and Offshores (OS). b) Analysis including AT, OS, and Alaska Residents (AR). Colour approximately represents population

size according to the key. Solid lines indicate migration rate  $> 1$  individual per generation and stippled lines indicate rates  $< 1$ , with thickness approximately representing migration rate. See Supplemental Tables S4 & S5 for parameter estimates. Ts stands for time segment and Ne stands for effective population size.

**Figure 1:**



Figure 2

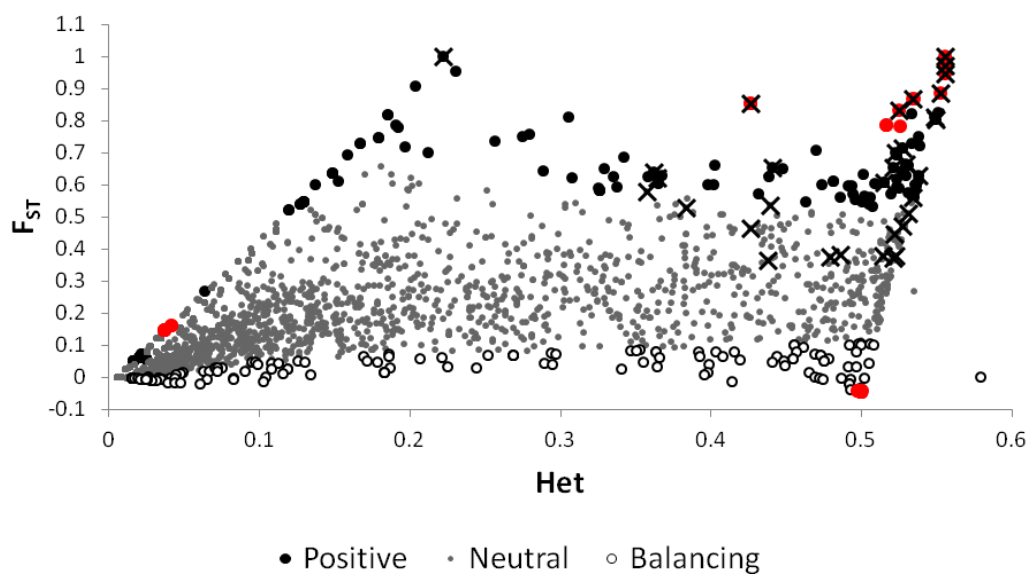


Figure 3

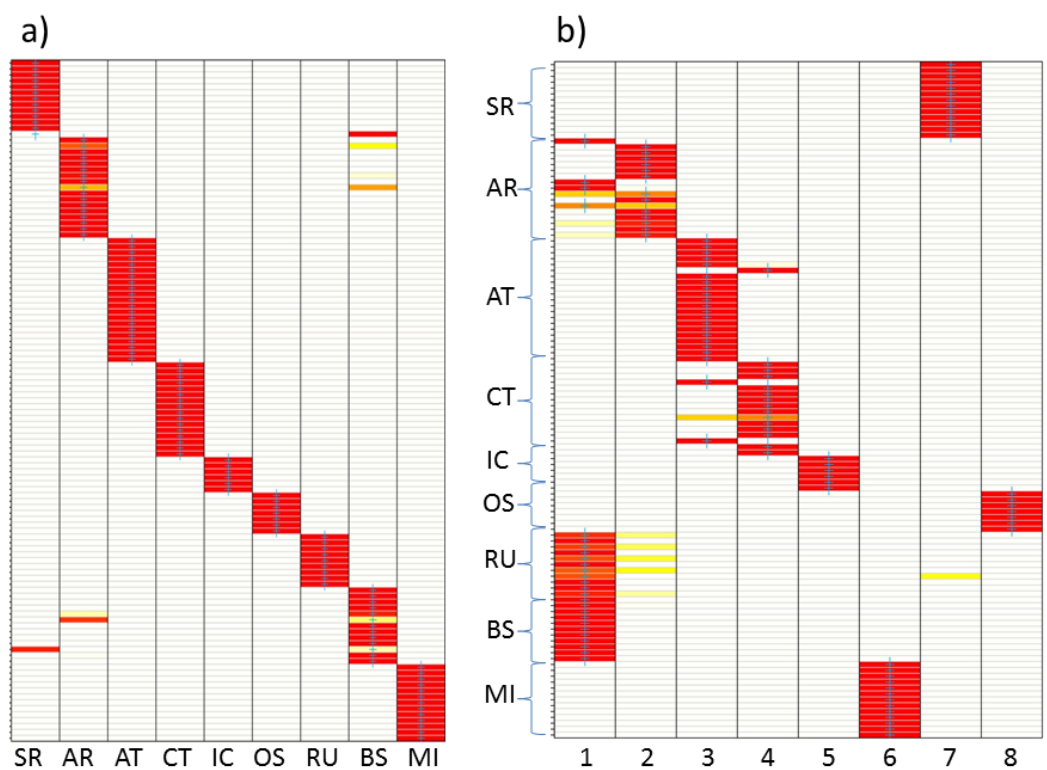


Figure 4

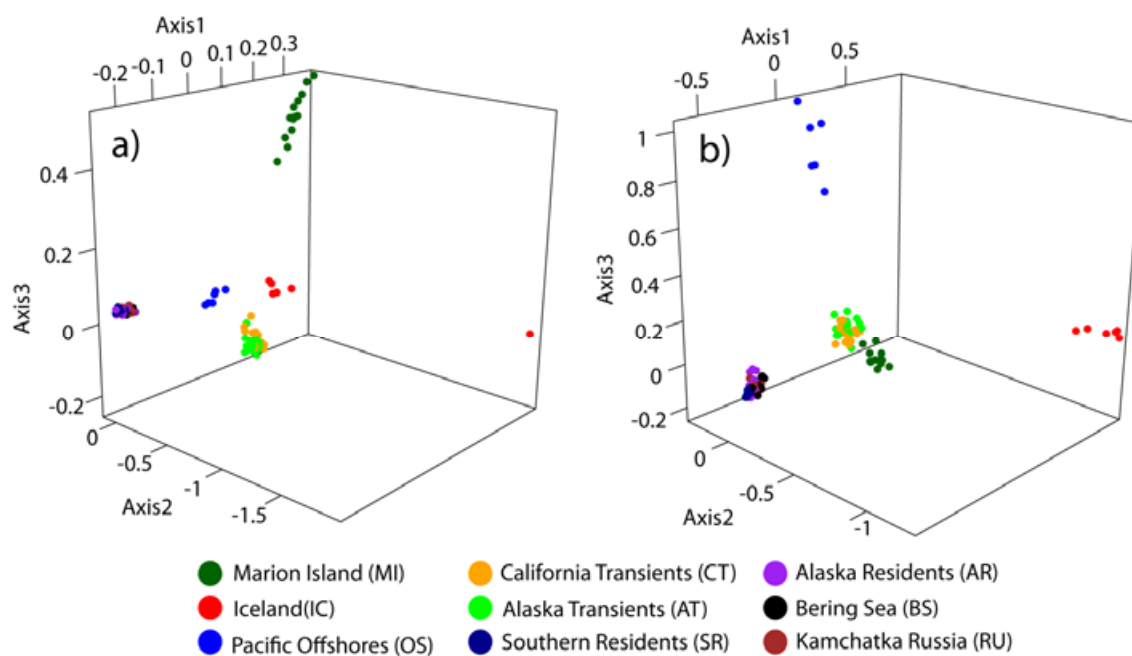


Figure 5

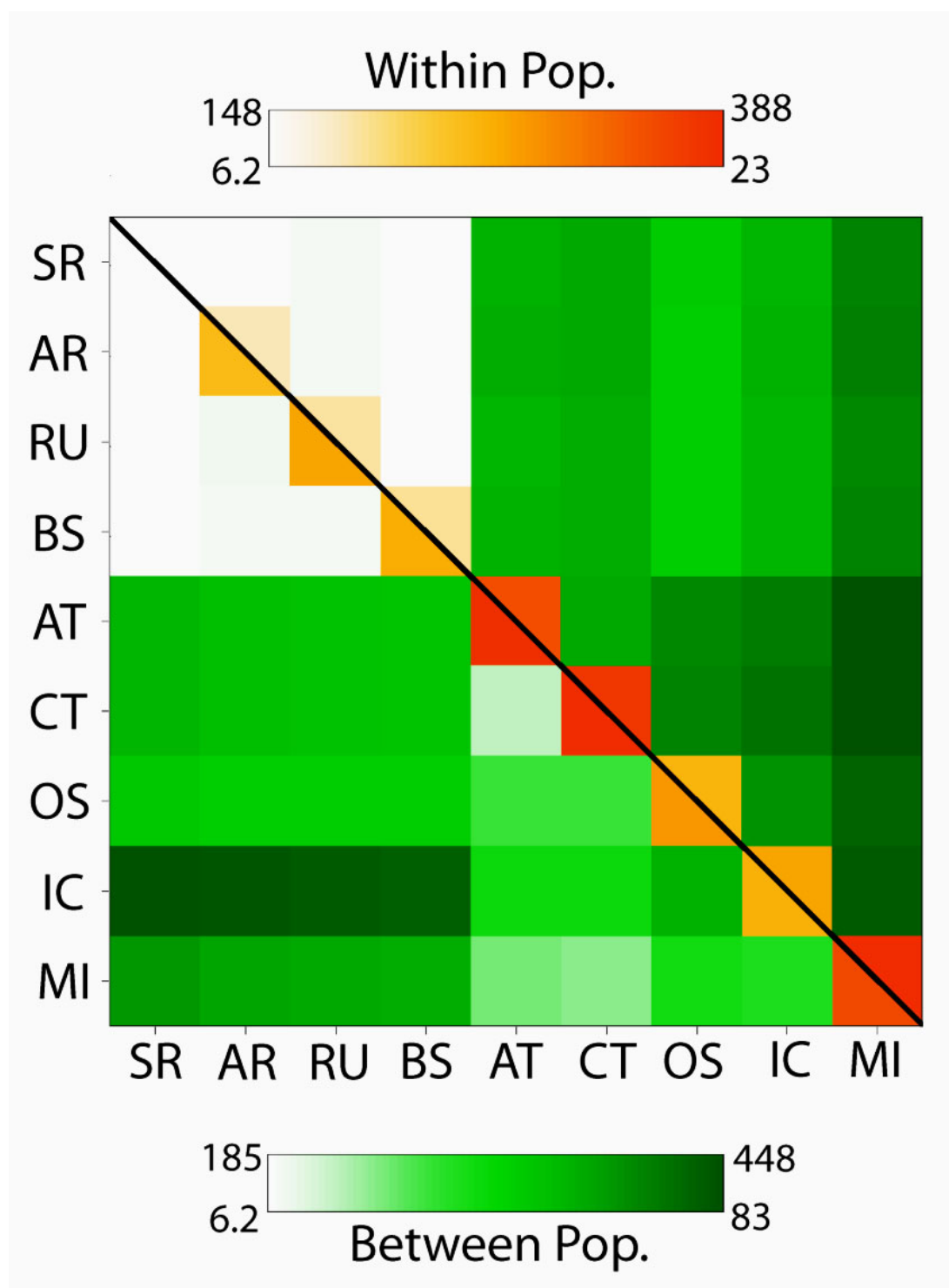


Figure 6

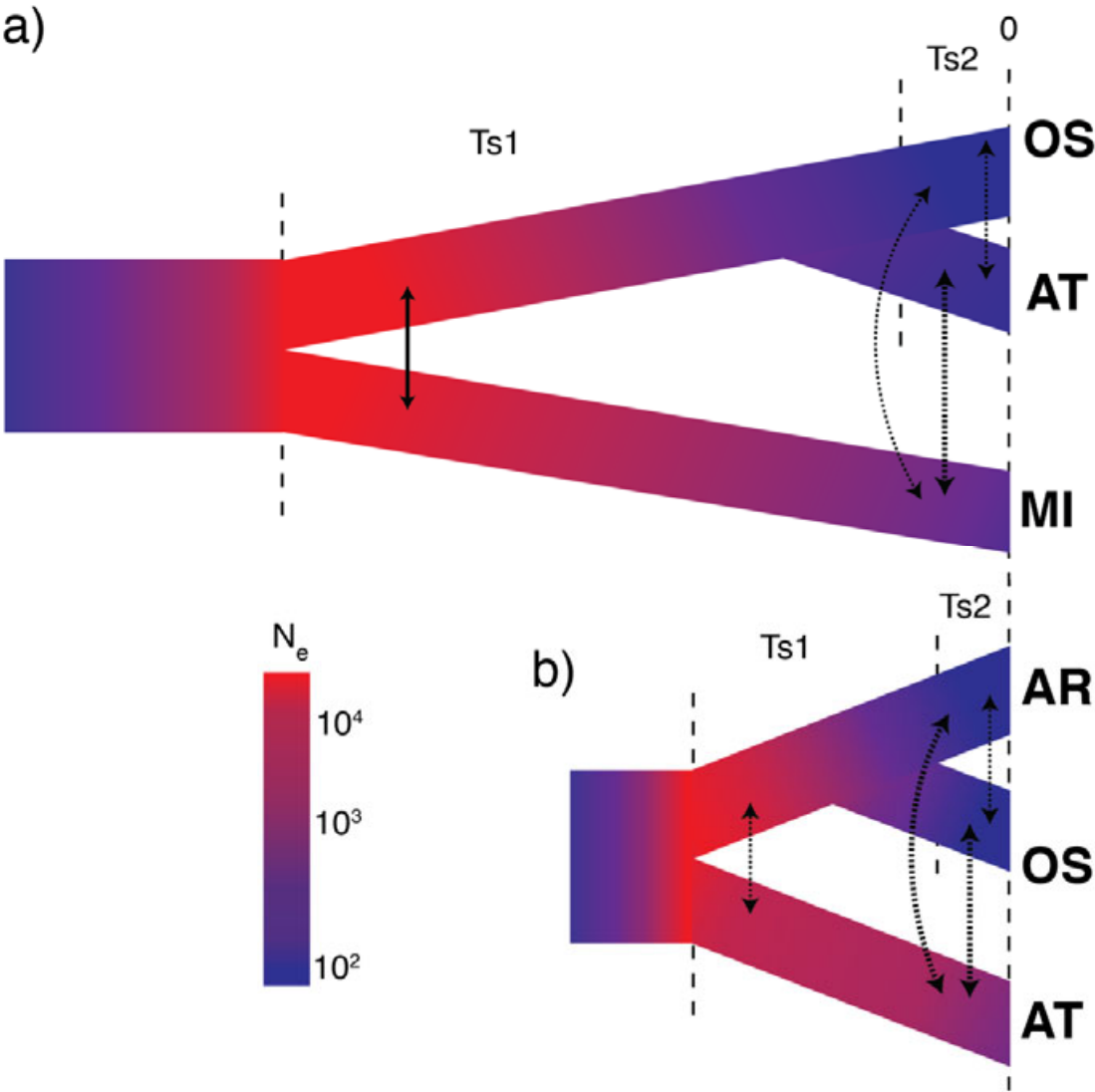


Table 1:  $F_{ST}$  comparisons among populations where  $N > 10$  for positive outlier loci (upper diagonal) and neutral loci (lower diagonal). All comparisons are highly significant ( $p < 0.001$ ). Abbreviations are as defined in Table 2.

|    | SR      | AR      | BS      | AT      | CT      | MI      |
|----|---------|---------|---------|---------|---------|---------|
| SR | 0       | 0.2475  | 0.21784 | 0.72847 | 0.73832 | 0.79715 |
| AR | 0.1632  | 0       | 0.05128 | 0.67336 | 0.6764  | 0.73618 |
| BS | 0.1363  | 0.04735 | 0       | 0.64762 | 0.64937 | 0.71197 |
| AT | 0.29719 | 0.26584 | 0.23356 | 0       | 0.05099 | 0.31034 |
| CT | 0.29339 | 0.25936 | 0.22429 | 0.0346  | 0       | 0.25045 |
| MI | 0.33442 | 0.30703 | 0.27442 | 0.17709 | 0.15502 | 0       |

Table 2: Number of fixed differences between populations. Sample size given as (N). Neutral loci on lower diagonal, outliers (for positive selection) on upper diagonal. Comparisons that show a significant number of fixed differences beyond that expected from

|         | SR | AR       | BS       | RU       | OS        | AT        | CT        | IC         | MI        |
|---------|----|----------|----------|----------|-----------|-----------|-----------|------------|-----------|
| SR (13) |    | <b>0</b> | <b>0</b> | <b>0</b> | <b>8*</b> | <b>7*</b> | <b>5*</b> | <b>15*</b> | <b>9*</b> |
| AR (17) | 0  |          | <b>0</b> | <b>0</b> | <b>5*</b> | <b>3</b>  | <b>2</b>  | <b>12*</b> | <b>6*</b> |
| BS (13) | 0  | 0        |          | <b>0</b> | <b>5*</b> | <b>3</b>  | <b>2</b>  | <b>9*</b>  | <b>5*</b> |
| RU (9)  | 0  | 0        | 0        |          | <b>4*</b> | <b>3</b>  | <b>2</b>  | <b>7*</b>  | <b>4*</b> |

sampling effects by chance are shown with an asterisk. SR= southern (Washington State) residents, AR = Alaskan residents, BS = the Bering Sea residents, RU = the Kamchatka, Russia residents, AT = Alaskan transients, CT = Californian transients, IC = Iceland, MI = Marion Island.



|         |    |    |   |   |   |   |   |    |   |
|---------|----|----|---|---|---|---|---|----|---|
| OS (7)  | 5* | 1  | 0 | 0 |   | 1 | 1 | 7* | 2 |
| AT (21) | 0  | 0  | 0 | 0 | 0 |   | 0 | 1  | 0 |
| CT (16) | 0  | 0  | 0 | 0 | 0 | 0 |   | 0  | 0 |
| IC (6)  | 6* | 4* | 3 | 3 | 2 | 0 | 0 |    | 0 |
| MI (13) | 2  | 0  | 0 | 0 | 0 | 0 | 0 | 0  |   |