Genome-culture coevolution promotes rapid divergence of killer whale ecotypes

Andrew D. Foote1,2,3,*, Nagarjun Vijay1,*, Maria C. Ávila-Arcos2,4, Robin W. Baird5, John W. Durban6, Matteo Fumagalli7, Richard A. Gibbs8, M. Bradley Hanson9, Thorfinn S. Korneliussen2, Michael D. Martin2, Kelly M. Robertson5, Vitor C. Sousa3, Filipe G. Vieira2, Tomáš Vinař10, Paul Wade11, Kim C. Worley8, Laurent Excoffier3, Phillip A. Morin6, M. Thomas P. Gilbert2,12 & Jochen B.W. Wolf1,13,14

Analysing population genomic data from killer whale ecotypes, which we estimate have globally radiated within less than 250,000 years, we show that genetic structuring including the segregation of potentially functional alleles is associated with socially inherited ecological niche. Reconstruction of ancestral demographic history revealed bottlenecks during founder events, likely promoting ecological divergence and genetic drift resulting in a wide range of genome-wide differentiation between pairs of allopatric and sympatric ecotypes. Functional enrichment analyses provided evidence for regional genomic divergence associated with habitat, dietary preferences and post-zygotic reproductive isolation. Our findings are consistent with expansion of small founder groups into novel niches by an initial plastic behavioural response, perpetuated by social learning imposing an altered natural selection regime. The study constitutes an important step towards an understanding of the complex interaction between demographic history, culture, ecological adaptation and evolution at the genomic level.
The interplay between ecology, culture and evolution at the level of the genome remains poorly understood. The ability to adapt to novel ecological conditions through behavioural plasticity is thought to be able to buffer natural selection pressures and promote rapid colonization of novel niches. However, by perpetuating exposure to a novel environment, stable cultural transmission of behaviour can also provide an opportunity for natural selection to act on adaptive genomic variation. Examples of genomic adaptation in humans during the period of recent ecological and cultural diversification and consequent demographic expansion are well illustrated.

For example, the Inuit of Greenland descend from a small founder population that split from an East Asian source population and successfully colonized the extreme climatic conditions of the Arctic environment through culturally transmitted methods of hunting marine mammals and genetic adaptation to a cold climate and hypoglycaemic lipid-rich diet. However, our understanding of the complex interaction between ecology, culture, adaptation and reproductive isolation at a genome-wide level has long suffered from deficiency of genome-wide data, and, conceptually, from the almost-exclusive focus on these processes in humans and thus a lack of comparative data from other species.

Killer whales (Orcinus Orca) are the largest species in the dolphin family (Delphinidae) and, together with humans, are one of the most cosmopolitan mammals, being found in all ocean basins and distributed from the Antarctic to the Arctic. This top marine predator consumes a diverse range of prey species, including birds, fish, mammals and reptiles. However, in several locations killer whales have evolved into specialized ecotypes, with hunting strategies adapted to exploit narrow ecological niches (see Supplementary Notes for more detailed information on the natural history of killer whale ecotypes). For example, in the North Pacific, two sympatic ecotypes coexist in coastal waters: the mammal-eating (so-called ‘transient’) ecotype and fish-eating (so-called ‘resident’) ecotype. This ecotypic variation is stable among multiple subpopulations of the transient and resident ecotypes across the North Pacific that diverged from common ancestral matrilineages ~68 and 35 KYA, respectively. A highly stable matrilineal group structure and a long post-menopausal lifespan in killer whales is thought to facilitate the transfer of ecological and social knowledge from matriarchs to their kin, and thereby perpetuate the stability of ecotypic variation in killer whales. In the absence of a definitional consensus and for the purposes of investigating how cultural phenomena interact with genes, culture has been broadly defined as ‘information that is capable of affecting individuals’ behaviour, which they acquire from other individuals through teaching, imitation and other forms of social learning’. Several studies have argued that behavioural differences among killer whale ecotypes are examples of culture in this broader sense of the term. However, this behavioural variation among ecotypes likely results from ecological, genetic and cultural variation and the interaction between them, rather than a single process explaining all behavioural variance. Killer whales, therefore, offer a prime example of how behavioural innovation perpetuated by cultural transmission may have enabled access to novel ecological conditions with altered selection regimes, and thus provide an excellent study system for understanding the interaction between ecological and behavioural variation, and genome-level evolution.

**Results and Discussion**

**Whole-genome sequencing.** We generated whole-genome re-sequencing data of 48 individuals at low coverage and accessed high-coverage sequencing data from two more individuals to investigate patterns of genomic variation among killer whale ecotypes. The samples represent five distinct ecotypes that, based on phylogenetic analysis of mitochondrial genomes, include some of the oldest and youngest divergences within the species (Fig. 1). The dataset incorporated 10 individuals each of the transient and resident ecotypes that occur in sympatry in the North Pacific; and from Antarctic waters, 7 individuals of a large mammal-eating form (type B1), 11 individuals of a partially sympatric, smaller form which feeds on penguins (type B2), and 10 individuals of the smallest form of killer whale, which feeds on fish (type C) (Fig. 1). A total of 2,577 million reads uniquely mapped to the 2.4-Gbp killer whale reference genome (Supplementary Fig. 1) for which a chromosomal assembly was generated for this study, so that approximately 50% of the autosomal regions of each individual were sequenced at ≥2× coverage (Supplementary Table 1). Subsequent data analyses used methods that account for uncertainty in the assignments of genotypes, enabling accurate inferences to be drawn from low-pass next-generation sequencing data. Comparisons of estimated population genomic metrics such as genome-wide and per-site Fst indicated that estimates from our low-coverage data were highly consistent with published high-coverage restriction site associated DNA sequencing (RAD-seq) and single-nucleotide polymorphism (SNP)-typing data (Supplementary Fig. 2 and Supplementary Table 2) and thus confirmed the robustness of our estimates.

**Time to most recent common ancestor.** We estimated a time to most recent common ancestor (TMRCA) of ~126–227 KYA from the accumulation of derived mutations at third-codon positions for the most divergent killer whale lineages compared here (Supplementary Table 3) and based on the 95% highest posterior density interval of the mutation rate estimate. This equates to ~4,900–8,800 generations and indicates a rapid diversification over a timescale comparable to the diversification of modern humans. We caution that for these age estimates we rely on mutation rate estimates derived from interspecific comparisons among odontocetes, and that, therefore, these estimates of TMRCA are at best approximate. Further, the demographic history and any gene flow between ecotypes will have an influence on the sharing of derived mutations and hence this estimate. However, we do expect that the estimate will be within the correct order of magnitude. Our estimated TMRCA overlaps with a recent RAD marker study, which estimated a TMRCA of 189 KYA (only a point estimate was reported by these authors) scaling by a mutation rate 1.2 times higher than we have used here. The estimated TMRCA of a global data set of killer whales based on non-recombining mitochondrial genomes has been estimated at 220–530 KYA (ref. 11), older than our estimate based on nuclear genomes. Male-mediated gene flow has therefore continued after matrilineal lineages have diverged.

**Genetic differentiation and divergence.** Despite this recent shared ancestry, substantial genome-wide differentiation and divergence had accrued between all pairs of ecotypes included in this study (Fig. 2 and Supplementary Table 2). At K = 5 populations, a maximum-likelihood-based clustering algorithm unambiguously assigned all individuals to populations corresponding to ecotype (Fig. 2c), indicating that all ecotypes have been assortatively mating long enough to allow allele frequencies to drift apart. Pairwise genetic distances between individuals visualized as a tree indicate that segregating alleles are largely shared within an ecotype (Supplementary Fig. 3). Similarly, pairwise relatedness due to identity-by-descent, that is, genetic identity because of a recent common ancestor,
was high within each ecotype. While the three Antarctic ecotypes still showed signs of recent relatedness, no shared recent identity-by-descent ancestry was detected between Antarctic and Pacific types or between the sympatric resident and transient ecotypes (Supplementary Fig. 4). The greatest differentiation (Supplementary Table 2) as visualized in the maximum likelihood graph (Fig. 2a) and PCA plot (Fig. 2b) was between the allopatric Pacific and the Antarctic ecotypes, while differentiation among Antarctic ecotypes was much lower than between the two Pacific ecotypes. Thus, our sampled populations allowed us to investigate the accrual of genomic differentiation along points of a continuum, acting as a proxy of sampling at different stages of the speciation process. The accrual of genome-wide differentiation ($F_{ST} = 0.09$) between even the most recently diverged and partially sympatric ecotypes (Antarctic types B1 and B2) indicates that reproductive isolation quickly becomes established after the formation of new ecotypes. Thus, whole-genome resolution confirms that, even in sympatry, contemporary gene flow occurs almost exclusively among individuals of the same ecotype, allowing genomic differentiation to build up between ecotypes so that within an ocean basin ecological variation better predicted genetic structuring than geography.

**Ancient admixture.** To better understand and visualize the complexity of the ancestry of killer whale ecotypes, we reconstructed the genetic relationships among ecotypes in the form of a maximum likelihood graph (Fig. 2a), representing the degree of genetic drift and modelling both population splits and gene flow using the unified statistical framework implemented in TreeMix (ref. 24). The inferred migration edges were supported by the three-population ($\beta$) and D-statistic (ABBA-BABA) tests, which can provide clear evidence of admixture, even if the gene flow events occurred hundreds of generations ago. These population genomic methods test for asymmetry in the covariance of allele frequencies that indicate that the relationships among populations are not fully described by a simple bifurcating tree model.

The three approaches were consistent in inferring migration from source populations that share ancestry with the North Pacific resident and North Atlantic ecotype into the transient ecotype (Fig. 2a, Supplementary Figs 5 and 6, Supplementary Tables 4 and 5 and Supplementary Data 1). The genomes of transients are therefore partly derived from at least one population related to the Atlantic and resident ecotypes, (but not necessarily these populations, that is, the source could be an unsampled ‘ghost’ population). The asymmetrical two-dimensional (2D)-site frequency spectrum (SFS) also implies directional gene flow from a population ancestrally related to the residents into the transient ecotype (Fig. 2e). Given the extent of the inferred demographic bottlenecks during founder events (see section below), and the expected consequential shift in the
SFS	extsuperscript{28}, it seems likely that this admixture would have occurred during or after the founder bottleneck in the transient ecotype, otherwise it would be expected to be less correlated with the resident SFS. The sequencing of more populations is expected to shed further light on this episode of ancient admixture.

TreeMix, the three-population and D-statistic tests inferred that type B1 is admixed and derives from at least two populations related to both types B2 and C (Fig. 2a, Supplementary Tables 4 and 5 and Supplementary Data 1). Conducting D-statistic tests on proposed tree-like histories comprising combinations of 16 genome sequences that included the North Atlantic sequence, we found that Antarctic types B1 and B2 shared an excess of alleles with the three Northern Hemisphere ecotypes (Supplementary Figs 7 and 8). This shared ancestry component between Northern and Southern hemisphere ecotypes was not detected in type C, suggesting a relatively recent admixture event after type C split from the shared ancestor of types B1 and B2. Other signals of ancient admixture among populations were also detected and are reported in the tables in the Supplementary Material.

Demographic history. As our results suggest that killer whale ecotypes have diversified rapidly from a recent ancestor, and given the importance of the relationship between effective population size ($N_e$) and the rate of evolution	extsuperscript{29} we conducted analyses to reconstruct their demographic history. Applying the pairwise sequential Markovian coalescent (PSMC) approach	extsuperscript{21} to two high-coverage ($\geq 20 \times$) autosomal assemblies, a North Atlantic female and a North Pacific resident male, refining the methodological approach of a previous analysis of these genomes (Supplementary Figs 9 and 10), we recovered a similar demographic trajectory to that previously reported	extsuperscript{17} (Fig. 3a). The inference of the timing of these demographic declines is dependent on the assumed mutation rate ($\mu$); however, across a range of sensible estimates of $\mu$, the declines broadly fall within the Late Pleistocene	extsuperscript{17}. This was previously interpreted as evidence for demographically independent population declines in each ocean, driven by environmental change during the Weichselian glacial period	extsuperscript{17}. However, this inference assumes that each PSMC plot tracks the demographic history of a single unstructured panmictic population	extsuperscript{30}. The $y$ axis of the PSMC plot is an estimation of $N_e$ derived from the rate of coalescence between the two chromosomes of a diploid genome. However, in the presence of population structuring, regions of the two chromosomes will coalesce less frequently as their ancestry may derive from different demes or subpopulations; the rate of coalescence can thus be similar to a single population with large $N_e$. PSMC estimates of $N_e$ during population splits can therefore be greater than the sum of the effective sizes of the subpopulations, dependent on the number of subpopulations and the degree of connectivity (that is, cross-coalescence) between them	extsuperscript{25,30}. In fact, the results presented here and previously published	extsuperscript{17,19,22} indicate that throughout the Weichselian glacial period there were multiple population splits, both between and within ecotypes, including the splitting of the two lineages included in the PSMC analyses just at the point of the change in inferred $N_e$ (Fig. 3a and Supplementary Fig 11). Therefore, the

Figure 2 | Evolutionary relationships among killer whale ecotypes. (a) TreeMix maximum likelihood graph from whole-genome sequencing data, rooted with the ‘transient’ ecotype. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred along that branch. The scale bar shows 10 times the average s.e. of the entries in the sample covariance matrix. Migration edges inferred using TreeMix and supported by the $F_3$ statistic test are depicted as arrows coloured by migration weight. (b) PCA, both the first and second principal components were statistically significant ($P$-value < 0.001). (c) Ancestry proportions for each of the 48 individuals conditional on the number of genetic clusters ($K = 3$–5). (d-f) Joint site frequency spectra for pairwise comparisons illustrating the speciation continuum. Each entry in the matrix $c_{xy}$ corresponds to the probability of observing a SNP with frequency of derived allele $x$ in population 1 and $y$ in population 2. The colours represent the probability for each cell of the SFS, white cells correspond to a probability of zero. The analysis illustrates that the amount of genetic differentiation is greater between (d) Pacific and Antarctic populations and (e) the pair of Pacific types than (f) among the evolutionarily younger Antarctic types, which have highly correlated site frequency spectra.
From ~300,000 to 130,000 years BP during the Saalian glacial period, the inferred \( N_{eX} \) ranged from 0.58 to 0.79 of the \( N_{eA} \) (Fig. 3b), which lies within expectation for demographically stable mammalian species\(^3\). During the first part of the Weichselian glacial period, \( N_{eX} \) markedly declined, reaching a minimum at ~30,000–50,000 years BP. The timing of this bottleneck in \( N_{eX} \) overlaps with the stem age of the mitochondrial clade for this Atlantic population\(^1\), that is, consistent with almost all mitochondrial diversity being lost in this lineage during this period. Conversely, this is concurrent with the peak estimate of autosomal \( N_{eA} \) inferred by PSMC, and ratio of \( N_{eX}/N_{eA} \) falls to ~0.3 at this \( N_{eX} \) minimum and then recovers within 1,000 generations to >0.75 (Fig. 3b). Simulated demographic bottlenecks of several hundredfold reduction result in a disproportionate loss of \( N_{eX} \), attributed to the difference in the inheritance mode of each marker, and the ratio of \( N_{eX}/N_{eA} \) can reach less than 0.3 (ref. 31). Following the bottleneck, \( N_{eX} \) recovers more rapidly than \( N_{eA} \) and the ratio of \( N_{eX}/N_{eA} \) can exceed 0.75 during this recovery phase\(^3^1\). The concordance of the timing of the bottleneck in the X-chromosome and the stem age of the mitochondrial genome suggest an underlying demographic process, rather than a strong selection on the X-chromosome or some other factors driving the mutation rate\(^3^1\). A sex-biased process such as primarily male-mediated gene flow between demes could further influence the ratio of \( N_{eX}/N_{eA} \) (ref. 32).

To estimate the demographic histories from our population genomic data, we produced ‘stairway’ plots using composite likelihood estimations of theta (\( \theta \)) for different SNP frequency spectra associated with different epochs, which are then scaled by the mutation rate to estimate \( N_e \) for each epoch\(^3\). Using this method we reconstructed a demographic history from our population genomic data for the resident ecotype that was comparable to the PSMC plot from a single resident high-coverage genome, both methods identifying a decline in \( N_e \) starting at ~60 KYA (Fig. 3a,c). The stairway plots for the transient ecotype and type C showed the same pattern as the resident of a decline to a bottlenecked population with an \( N_e \) of <1,000 and a subsequent expansion (Fig. 3c). The bottlenecks did not occur simultaneously, as might be expected in response to a global environmental stressor during a glacial cycle, but instead were sequential (Fig. 3c). In each case, the timing of the demographic bottleneck overlapped with the previously estimated timing of the stem age of the mitochondrial genome clades containing each ecotype\(^1\). Thus, within the transient, resident and combined Antarctic ecotypes, both mitochondrial and nuclear lineages coalesce back to these respective bottleneck events, consistent with genetic isolation of small matrifocal founder groups from an ancestral source population, followed by the subsequent expansion and substructuring of a newly established ecotype.

Overall, the population genetic analyses of the whole-genome sequences above shed light on the ancestry of killer whale ecotypes in unprecedented detail, highlighting a complex tapestry of periods of isolation interspersed with episodic admixture events and strong demographic bottlenecks associated with the founder events that gave rise to the resident, transient and ancestral Antarctic ecotypes.

**Genome-wide landscape of genetic diversity and differentiation.** Demographic bottlenecks during population splits and founding events, followed by subsequent demographic and geographic expansion, can produce rapid shifts in allele frequencies between populations\(^3\). The high levels of genome-wide differentiation (\( F_{ST} \)) between killer whale ecotypes across all genomic regions (Fig. 4a,b) are consistent with strong genetic drift following demographic expansion from small founding groups.
Considering the low efficiency of selection in populations as small as the estimates presented here\textsuperscript{35}, in which founder populations have an estimated $N_e$ ranging from a few tens to hundreds, a genome-wide contribution of ecologically mediated divergent selection is neither necessary nor particularly likely to explain the observed shifts in allele frequencies in such a large number of loci. Consistent with this prediction, we find that differentiation is highest along the branches inferred by TreeMix to have experienced the most substantial genetic drift (Fig. 2a), that is, the branch to the ancestor of the Antarctic types and the branch to the resident ecotype (Supplementary Fig. 13). We therefore expect that only those beneficial alleles that have a strong favourable effect (that is, strength of selection ($s > 1/2N_e$)) would have an increased fixation probability because of selection within these founder populations.

Much of the heterogeneity in the differentiation landscape was shared among pairwise comparisons (Fig. 4c). Since diversity ($\pi$) was not associated with mutation rate $\mu$, as inferred from neutral substitution rate $dS$, ($r = -0.1$ to $-0.24$), the observed covariation in differentiation ($F_{ST}$), diversity ($\pi$) and absolute divergence ($D_{xy}$) between population pairs (Fig. 4c and Supplementary Fig. 14) is best explained by the shared local reduction of diversity by linked selection in the ancestral population\textsuperscript{36,37}. This leads us to conclude that overall the
landscape of genome-wide differentiation is a result of global genetic drift, regionally elevated by ancestral linked selection (for example, background selection or repeated selective sweeps shared among populations) independent of the evolutionary dynamics of the recently derived present-day ecotypes.

Genomic signatures of climate and diet adaptation. However, against this background of shared differentiation, there was evidence for genic divergence of putative functional relevance. The first targets of selection following ecotype diversification and the exploitation of new ecological niches are expected to be those that facilitate ecological specialization46. Once a certain level of reproductive isolation is reached, differences can accumulate in other (fast-evolving) genes involved in reproductive isolation effectively reducing hybrid mating38. Following this rationale, both individual gene associations and gene ontology (GO) enrichment analyses yielded several biological processes and candidate genes with putative functional roles in ecological specialization, local adaptation and reproductive isolation (Supplementary Tables 6–8). For example, in comparisons between ecotypes inhabiting the extreme cold of the Antarctic pack ice with ecotypes from the more temperate North Pacific (Fig. 1), we found the most significant enrichment in genes involved in adipose tissue development (GO:0006012, Fisher’s exact test: \( P = 0.0015 \)). Genes associated with adipose tissue development have previously been found to be evolving under positive selection in the polar bear39, suggesting a role for this process in rapid adaptation to a cold climate and lipid-rich diet.

Using the population branch statistic (PBS), which has strong power to detect recent natural selection40 and has allowed us to investigate allele changes along specific branches, we identified another candidate example where cold adaptation may play a role. The FAM83H gene showed a signature of selection (top 99.9% PBS values) and was found to contain four fixed non-synonymous substitutions derived in the Antarctic lineages based on the inferred ancestral state, which resulted in physicochemical changes including a hydrophobic side chain being replaced by a positively charged side chain. The keratin-associated protein encoded by the FAM83H gene is thought to be important for skin development and regulation through regulation of the filamentous state of keratin within cytoskeletal networks in epithelial cells, determining processes such as cell migration and polarization41. Skin regeneration is thought to be constrained in killer whales while inhabiting the cold waters around Antarctica because of the high cost of heat loss, and is thought to underlie rapid round-trip movements to warmer subtropical waters by Antarctic ecotypes42. The balance between skin regeneration and thermal regulation in Antarctic waters could be a major selective force requiring both behavioural42 and genomic adaptation (Supplementary Fig. 15).

Genes encoding proteins associated with dietary variation also showed a signature of selection (top 99.9% PBS values). For example, the carboxylesterase 2 (CES2) gene encodes the major intestinal enzyme and has a role in fatty acyl and cholesterol ester metabolism in humans and other mammals43. Two exons of the CES2 gene had among the top 99.9 percentile PBS values because of changes in allele frequencies (including two fixed non-synonymous amino-acid changes) along the branch to the Antarctic types (Supplementary Fig. 16). Similarly, genes in the top 99.9 percentile PBS values were enriched for carboxylic ester hydrolase activity (GO:00052689, Fisher’s exact test: \( P < 0.0001 \)) in the fish-eating resident ecotype. Biological processes enriched in the resident killer whale included digestive tract morphogenesis (GO:0048546, Fisher’s exact test: \( P = 0.0022 \)), and gastrulation with mouth forming second (GO:0001702, Fisher’s exact test: \( P = 0.0024 \)): associated with the formation of the three primary germ layers of the digestive system during embryonic development. These results overlapped with enriched GO terms identified by a previously published RAD-seq study, despite the relatively sparse sampling of 3,281 SNPs in that study19. Enrichment of these GO terms was largely driven by differentiation in a single exon in the GATA4 gene, which included a fixed non-synonymous substitution, sequenced in both studies.

Signatures of selection along branches leading to the two predominantly mammal-eating ecotypes included in this study, the North Pacific transient and Antarctic type B1, were found in genes that play a key role in the methionine cycle (Fig. 5). Methionine is an essential amino acid that has to be obtained through dietary intake, and is converted through trans-sulfurcration to cysteine via intermediate steps of catalysis to homocysteine44. Any excess homocysteine is re-methylated to methionine44. Diets with different protein contents, such as between killer whale ecotypes, will differ in their content of methionine, and the enzymatic cofactors involved in the metabolism of methionine and homocysteine, which include folate, vitamins B6 and B12 (ref. 44; hence why vegetarians often take vitamin B12 supplements). While different genes and different biological processes showed a signature of selection in each of these two mammal-eating ecotypes (Fig. 5), in both cases the candidate genes and processes were associated with the regulation of methionine metabolism, which results in the generation of cysteine. Successful hunting of mammal prey by killer whales would provide a sudden and rich source of dietary methionine. This fluctuating intake of protein may place more of a selective pressure on the regulation of the metabolism of methionine than does the consumption of fish by piscivorous ecotypes.

Rapid evolution of reproductive proteins. High (top 99.9 percentile) PBS values, largely driven by fixed non-synonymous amino-acid substitutions, were further estimated for several genes that encode proteins associated with reproductive function, including tests development, regulation of spermatogenesis, spermatocyte development and survival, and initiating the acrosome reaction of the sperm (for example, PKDREJ, RXFP2, C9orf24, SPEF1, TSSK4, DHH and MMEL1). Reproductive proteins such as PKDREJ, in which we found two fixed non-synonymous substitutions derived in the ‘resident’ ecotype, are known to diverge rapidly across taxa, and because of their functional role in fertilization are emerging as candidates for the post-zygotic component of the speciation process38.

Conclusions

Overall, our results indicate that the processes underlying genomic divergence among killer whale ecotypes reflect those described in humans in several respects. Behavioural adaptation has facilitated the colonization of novel habitats and ecological niches. Founder effects and rapid formation of reproductive isolation, followed by population expansion, have promoted genome-wide shifts in the frequency of alternative alleles in different ecotypes due to genetic drift. Demographic changes during founder events and subsequent expansions can also influence cultural diversity45,46, and may have had a role in reducing within-ecotype cultural diversity and promoting cultural differentiation between ecotypes. As with studies on modern humans, it is difficult to demonstrate a causal association between cultural differences and selection on specific genes1; however, our findings of divergence in genes with putative functional association with diet, climate and reproductive isolation broadly imply an interaction between genetically and culturally heritable
Methionine is an essential amino acid that is obtained through dietary protein. The methionine cycle feeds into the folate cycle, and both are part of the complex and interacting network of pathways that encompass one-carbon metabolism. Methionine adenyltransferase (MAT) then generates S-adenosylmethionine (SAM), which is then demethylated by the methyltransferases DNMT1 (ref. 68) and TRMT10A (ref. 69) to methylate DNA and RNA, respectively, and to form S-adenosylhomocysteine (SAH). SAH is converted to homocysteine, which is then catalysed by cystathionine β-synthase, an enzyme encoded by the CBS gene, to cystathionine, which is then converted to cysteine. Any excess homocysteine is re-methylated to methionine. Adapted and simplified from KEGG pathway map 00270. Genes that play a role in this pathway and with a signature of selection (top 99.9% PBS values) in either of the predominantly mammal-eating ecotypes (the transient or type B1) are in red boxes. The gene encoding the carbonic anhydrase CA5A, which plays a role in the larger one-carbon pathway, also had a signature of selection (top 99.99 PBS value) in type B1.

(b) Boxes indicate GO terms associated with the biological processes within the methionine cycle that were enriched in genes with the top 99.9 (transients; blue) and 99.99 (type B1; purple) percentile PBS values.

(c) Evolutionary trees underlying the signal of selection. Population-specific allele frequency changes are indicated by the FST-based (PBS values) branch lengths in these genes (red), which are overlaid on the genome-wide average branch lengths (black). The high (top 99.9-99.99%) PBS values indicate substantial changes in allele frequencies along the branches to the mammal-eating ecotypes. Differentiation in genes DNMT1, TRMT10A and CA5A was greatest between the mammal-eating type B1 and the fish-eating type C among the Antarctic types. (Photo credit of ‘dietary protein’: Robert Pitman, SWFSC.)
evolutionary changes in killer whale ecotypes. Given these findings, the almost-exclusive focus on humans by studies of the interaction of culture and genes should be expanded, and exploration of culture–genome coevolution models in suitable non-human animal systems encouraged.

**Methods**

**Sample collection.** Skin biopsies from free-ranging killer whales were collected using project-specific dart, concurrent with the collection of photographic, photogrammetry and behavioural data, allowing sampled individuals to be binned to ecotype/morphotype *a priori* to genetic analyses. Most samples were selected from separate collection dates and identified groups (when known) to minimize chances of collecting close relatives or replicative individuals. The sample set included 10 *resident*, 10 *transient*, 7 type B1, 11 type B2 and 10 type C killer whales.

**Genomic DNA library building and sequencing.** DNA was extracted using a variety of common extraction methods as per ref. 11. Genomic DNA was then sheared to an average size of ~150–200 bp using a Diagenode Bioruptor NGS. Illumina sequencing libraries were built on the sheared DNA extracts using NEBNext (Ipswich, MA, USA) DNA Sample Prep Master Mix Set 1 following Meyer and Kircher22. Libraries were subsequently index-amplified for 15 cycles using NEBNext High Quality Master Mix (Finzymes) in 50–110 µl reactions following the manufacturer’s guidelines. The libraries were then purified using the MiniElute PCR purification kit (Qiagen, Hilden, Germany). The DNA concentration of the libraries was measured using a 2100 Bioanalyzer (Agilent Technologies, CA, USA); these were then equimolarly pooled by ecotype and each ecotype pool was sequenced across five lanes of an Illumina HiSeq 2000 platform using single-read 100-bp chemistry (that is, a total of 25 lanes).

**Read trimming and mapping.** A high-quality, 2,249-Mb reference killer whale genome assembly (Orcas1.1, GenBank: ANO1L000000002.2, contig N50 size of 70.3 kb, scaffold N50 size of 12.7 Mb)16 was used as a mapping reference. For the purpose of this study, the genome was masked for repetitive elements using RepeatMasker48 and the Cetartiodactyl repeat library from Repbase49. Repetitive elements constitute 41.32% of the killer whale reference genome (926,443,262 sites). A further 80,599 sites were identified as mitochondrial DNA transposed to the nuclear genome (numts) and were masked accordingly. The final assembly was then indexed using BWA v.0.5.9 (ref. 50) to serve as the reference for read-mapping.

Illumina HiSeq 2000 reads from each individual were processed with AdapterRemoval51 to trim residual adapter sequence contamination and to remove adapter dimer sequences as well as low-quality stretches at 3′ ends. Filtered reads >30 bp were then mapped using Burrows-Wheeler aligner (BWA), requiring a mapping quality greater than 30. Clonal reads were collapsed using the random function of the SAMtools (v. 0.1.18) suite52. Ambiguously mapped reads were also filtered out using SAMTools. Consensus sequences were then reconstructed in Binary sequence/Alignment Map file format. To ensure that all repeat regions, which have the potential to bias population genetic inference, were removed, the Ti/Tv ratio of 3.8 at third-codon positions within our killer whale data set was used to predict the time taken to accumulate 124 derived transversions and 465 transitions at third-codon positions that we had inferred to be derived in type B1 since splitting from a shared ancestor with the resident ecotype.

**Admixture analysis.** An individual-based assignment test was performed to determine whether the ecotypes to which each individual had been assigned *a priori*, based on observed behaviour and/or morphological characteristics at the time of sampling, represented discrete gene pools in Hardy–Weinberg equilibrium. Since the 48 genomes generated for this study had an average sequencing depth of 2×, genotypes can only be called with very high uncertainty. Therefore, NGSadmix23, a maximum likelihood method that bases its inference on genotype likelihoods (GLs) and in doing so takes into account the uncertainty in the called genotypes that is inherently present in low-depth sequencing data, was employed. The method has been demonstrated, using simulations and publicly available sequencing data, to have great accuracy even for very low-depth data of less than twofold mean depth53. GLs were estimated using the SAMtools method52 implemented in the software package ANGSD54. NGSadmix was run with the number of ancestral populations *K* set to 3–5. For each of these *K* values, NGSadmix was re-run multiple times with different seeds in order to ensure convergence. Sites were further filtered to include only autosomal regions covered in at least 40 individuals, and removing sites with a minor allele frequency below 5% estimated from the GLs, resulting in the analyses being based on 603,519 variant sites. The highest likelihood solutions can be seen in Fig. 2c.

**Principal Component Analysis.** Assignment of individuals to ecotype, and structuring among ecotypes, was further investigated using the Principal Component Analysis (PCA), implemented in the ngstools suite54, taking genotype uncertainty into account55. Brieﬂy, the covariance matrix between individuals, computed as proposed in ref. 56, is approximated by weighting each genotype for its posterior probability, the latter computed using ANGSD as described in ref. 18. The eigenvectors from the covariance matrix were generated with the *R* function ‘eigen’, and significance was determined with a Tracy–Widom test to evaluate the statistical significance of each principal component identified by the PCA. The PCA was plotted using an in-house R script (available at https://github.com/mifumagalli/ngsPopGen/tree/master/scripts).

**Tests for ancestral admixture.** The genetic relationships among ecotypes were further reconstructed in the form of a maximum likelihood graph representing the degree of genetic drift generated using TreeMix35. TreeMix estimates a bifurcating maximum likelihood tree using population allele frequencies as input data and calculates transversion and transition mutation rates. A likelihood-based model implementation using a Gaussian approximation. The branches of this tree represent the relationship between populations based on the majority of alleles. Migration edges are then fitted between populations that are a poor fit to the tree model, and in which the exchange of alleles is inferred. The addition of migration edges between populations is undertaken in stepwise iterations to maximize the likelihood, until no further increase in statistical significance is achieved. The directionality of gene flow along migration edges is inferred from asymmetries in a covariance matrix of allele frequencies relative to an ancestral population as implied from the maximum likelihood tree of the transient ecotype as the root, with blocks of 1,000 SNPs (corresponding to approximately several hundred kilobases) to account for linkage disequilibrium among sites. The graphs are presented in
Estimating population genomic metrics. Measures of genetic differentiation, divergence and diversity were estimated using methods specifically designed for low-coverage sequencing data. Using a Maximum-Likelihood-based approach previously proposed and using the bottleneck dolphin genome to determine the ancestral state of each site, the unfolded SFS was estimated jointly for all individuals in a population for sites sequenced in five or more individuals in each population. Using this ML estimate of the SFS as a prior in an Empirical Bayes approach, the posterior probability of all possible allele frequencies at each site was computed. For these quantities, the expectations of the number of variable sites and fixed differences between lineages were estimated as the sum across sites of the probability of each site to be variable as previously proposed. Finally, the posterior expectation of the sample allele frequencies was calculated as the basis for further analysis of genetic variation within and between lineages.

We estimated the probability of a site being variable (Pvar) as input data. We also estimated the frequency of the most ancestral allele at each variable site which was used to determine the frequencies of the other alleles. The population-specific allele frequencies were estimated with GATK. Population-specific allele frequencies were estimated for sites covered by at least five individuals in each population in windows of size 50, 100 and 200 kb (Supplementary Table 9). This method does require genotype calls as an input, and is therefore susceptible to errors associated with genotype calling from low-coverage data. However, since TreeMix works on population allele frequencies, not genotypes, it was possible to determine the frequencies of the most common alleles with high confidence. The topology is comparable to output from other approaches applied here that do account for genotype uncertainty, providing confidence in the result.

Demographic reconstruction. PSMC analysis was performed on the high-coverage diploid autosomal genome sequences of two individuals to investigate changes in effective population (Ne) size. The PSMC model estimates the TMCA of segmental blocks of the genome and uses information from the rates of the two high-coverage genomes. Changes in Ne are inferred from the slopes of the past demographic changes of a population. The method has been validated by its successful reconstructions of demographic histories using simulated data and genome sequences from modern human populations.

A more detailed account of PSMC analyses, including site reanalyses and re-estimation of maximum-likelihood analysis of low-coverage genome assemblies, is presented in the Supplementary Figures and Notes. A PSMC plot of the Atlantic genome subsampled to 20 x and a North Pacific resident genome (20 x) was estimated, scaled to the autosomal mutation rate of 2.34 × 10^-8 substitutions per nucleotide per generation and is presented in Fig. 3a. A PSMC plot of the autosomal regions of the North Atlantic female killer whale at a coverage of 50 x was scaled to the autosomal mutation rate (0.24 x 10^-8) substitutions per nucleotide per generation, as used in our estimation of TMCA (see above) and was compared with a plot of the diploid X-chromosome, scaled to real time as per ref. 21 in which the neutral mutation rate of the X-chromosome was derived as \( \mu = \frac{2}{X+1} \), assuming a ratio of male-to-female mutation rate of \( a = 2 \) (ref. 61). This gave us an estimated \( \mu = 2.08 \times 10^{-8} \) substitutions per nucleotide per generation. The plot is presented in Fig. 3b. We also re-estimated the PSMC plot for the X-chromosome using different mutation rates to investigate which rate would produce PSMC plots with inferred concurrent declines in Ne in autosomes and X-chromosome. We found that \( \mu = 1.00 \times 10^{-8} \) substitutions per nucleotide per generation would be needed to synchronize the inferred demographic changes in these two markers (Supplementary Fig. 17). This would require the male-to-female mutation rate (a) to be orders of magnitude higher, making it seemingly biologically unrealistic.

The inference of demographic history from population genomic data by the Stairway Plot method provided a somewhat comparable result to reconstruction from a single 20 x genome by PSMC with respect to the time of onset and magnitude of a demographic decline inferred by both methods. The Stairway Plot method inferred a sudden drop in Ne, whereas a more gradual decline was inferred by PSMC, consistent with simulations showing that PSMC can infer abrupt changes in Ne as gradual changes. The sudden change in estimated effective population size is in the Stewartia case because of the step-like inference of the multi-epoch method, in which epochs coincide with coalescent events. Therefore, the Method was then applied to the site frequency spectra from a single 20 x genome by PSMC with respect to the time of onset and magnitude of a demographic decline inferred by both methods. The method was applied to the site frequency spectra of the transient type C. The results are shown in Fig. 3c, and in each case the Stairway Plot method inferred a sudden and dramatic decline in Ne. Therefore, and as inferred from the PSRC analysis of the low-coverage genomes, whereas PSMC did not infer an expansion within our cutoff point of 10,000 years ago, but did infer a more gradual expansion during the Holocene. It should be kept in mind that the PSRC plot is based on data from a single individual and so will be more susceptible to the effects of the single individual's history. Ne continues to split into multiple discrete populations. In contrast, the Stairway Plot is based on population genomic data and will track the change in Ne across all the sampled resident populations after they have split.

The inference of demographic history from population genomic data by the PSRC method provided a somewhat comparable result to reconstruction from a single 20 x genome by PSMC with respect to the time of onset and magnitude of a demographic decline inferred by both methods. The inference of demographic history from population genomic data by the PSRC method provided a somewhat comparable result to reconstruction from a single 20 x genome by PSMC with respect to the time of onset and magnitude of a demographic decline inferred by both methods. The PSRC method inferred a sudden drop in Ne, whereas a more gradual decline was inferred by PSMC, consistent with simulations showing that PSMC can infer abrupt changes in Ne as gradual changes. The sudden change in estimated effective population size is in the Stewartia case because of the step-like inference of the multi-epoch method, in which epochs coincide with coalescent events. Therefore, the PSRC method inferred a sudden and dramatic decline in Ne. Therefore, and as inferred from the PSRC analysis of the low-coverage genomes, whereas PSMC did not infer an expansion within our cutoff point of 10,000 years ago, but did infer a more gradual expansion during the Holocene. It should be kept in mind that the PSRC plot is based on data from a single individual and so will be more susceptible to the effects of the single individual's history. Ne continues to split into multiple discrete populations. In contrast, the Stairway Plot is based on population genomic data and will track the change in Ne across all the sampled resident populations after they have split.

The inference of demographic history from population genomic data by the Stairway Plot method provided a somewhat comparable result to reconstruction from a single 20 x genome by PSMC with respect to the time of onset and magnitude of a demographic decline inferred by both methods.
Inferring putatively functional allele shifts because of selection. Shifts in allele frequencies can occur because of selection, but differences in allele frequencies can also accumulate between populations because of drift. To infer shifts in allele frequencies potentially due to selection, we considered the 1% of SNPs with the highest \( F_{ST} \) values from each pairwise comparison between killer whale ecotypes. However, as \( F_{ST} \) is dependent on the underlying diversity of the locus, even extreme outlier loci can be due to genetic drift alone. We, therefore, additionally looked for over-representation of the top 1% outliers in different categories, for example, exons, 25-kb flanking regions (potential regulatory regions), introns and intergenic regions, at different \( F_{ST} \) bins using a \( \chi^2 \)-test. Residuals are expected to be normally distributed and indicate statistical significance of over- and under-representation of specific categories. The significant threshold was subjected to Bonferroni correction. The top 100 outliers in exons from each pairwise comparison were then used for GO over-representation analysis\(^6\) to identify enrichment due to diet (in mammal- versus fish-eating ecotypes) and climate (in Antarctic versus Pacific).

To more robustly infer whether genetic changes in exons were associated with ecotype divergence due to selection, we applied the PBS\(^2\). The PBS has the power to detect (even incomplete) selective sweeps over short divergence times, making it relevant for the scenario we are investigating in this study. We therefore estimated the PBS for 50-kb sliding windows shifting in 10-kb increments (approximating to a window size of 10 SNPs) to detect regions of high differentiation potentially due to selective sweeps. We followed the approach of Yi et al.\(^5\) and used the classical transformation by Cavalli-Sforza\(^3\): \( T = -\log (1 - F_{ST}) \) to estimate the order of magnitude of the population divergence time \( T \) in units scaled by the population size. For each 50-kb window, we calculated this value between each pair of ecotypes. The length of the branch leading to the one ecotype since the divergence from a recent ancestor (for example, in the equation given the length of the branch to type B1 since diverging from types B2 and C) is then obtained as

\[
PBS = \frac{F_{ST} - 1 + \sqrt{F_{ST} - 1 + 4F_{ST}}}{2}
\]

These window-based PBS values represent the amount of allele frequency change at a given 50-kb genomic region in the history of this ecotype (since its divergence from the other two populations)\(^6\). To further narrow down the target of selection, the PBS was estimated for exons farther ref. 5 and compared with genome-wide values to identify whether they were in the top 99.9 percentile (for the branches to the resident, transient and ancestral Antarctic ecotypes) and the top 99.99 percentile for the shorter branches leading to each Antarctic ecotype from their most recent common shared ancestor. We further filtered outliers to just include exons that encoded non-synonymous amino-acid substitutions and searched the database GeneCards\(^6\) for functions of the encoded proteins to identify potential targets of natural selection due to ecological differences. Supplementary Table 10 contains a list of outlier loci and the corresponding PBS value for each branch. Supplementary Data 2 contains details of fixed non-synonymous changes in the exons of protein-coding genes, including the per-individual and per-population read counts for each allele.

References

1. Laland, K. N., Odling-Smee, J. & Myles, S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat. Rev. Genet.* **11**, 137–148 (2010).
2. Wang, E. T., Kodama, G., Baldis, P. & Moyzis, R. K. Global landscape of recent gene flow and population divergence time. *Proc. Natl Acad. Sci. USA* **103**, 135–140 (2006).
3. Moura, A. E. et al. Phylogenomics of the killer whale indicates ecotype evolution in sympathy involving both selection and drift. *Mol. Biol. Evol.* **23**, 3179–3192 (2014).
4. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
5. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–720 (2010).
6. Patterson, N. et al. Ancient admixture in human history. *Genetics** **192**, 1065–1093 (2012).
7. Sousa, V. & Hey, J. Understanding the origin of species with genome-scale data: modelling gene flow. *Nat. Rev. Genet.* **14**, 404–414 (2013).
8. Math, G. T., Czabarka, E., Murvai, J. & Sherry, S. T. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics** **166**, 351–372 (2004).
9. Lanfear, R., Kokko, H. & Eyre-Walker, A. Population size and the rate of evolution. *Trends Ecol. Evol.* **29**, 31–41 (2014).
10. Rendell, L. & Whitehead, H. Culture in whales and dolphins. *Behav. Brain Sci.* **24**, 309–324 (2001).
11. Laland, K. N. & Janik, V. M. The animal cultures debate. *Trends Ecol. Evol.* **21**, 542–547 (2006).
12. Forster, A. D. et al. Convergent evolution of marine mammal genomes. *Nat. Genet.* **47**, 272–275 (2015).
13. Moura, A. E. et al. Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the Last Glacial Maximum. *Mol. Biol. Evol.* **31**, 1121–1131 (2014).
14. Nielsen, R., Korneliusen, T., Albrechtsen, A., Li, Y. & Wang, J. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS ONE* **7**, e37538 (2012).
15. Moura, A. E. et al. Population genomics of the killer whale indicates ecotype evolution in sympathy involving both selection and drift. *Mol. Biol. Evol.* **23**, 3179–3192 (2014).
16. Durban, J. W., Fearnbach, H., Burrows, D. G., Ylitalo, G. M. & Pitman, R. L. Antarctic killer whales make rapid, round-trip movements to subtropical waters: evidence for physiological maintenance migrations? *Biol. Lett.* **8**, 274–277 (2012).
17. Sato, T. & Hosokawa, M. The mammalian carboxylesterases: from molecules to functions. *Ann. Rev. Pharmacol. Toxicol.* **38**, 257–288 (1998).
44. Finkelstein, J. D. Methionine metabolism in mammals. *J. Nutr. Biochem.* 1, 228–237 (1990).
45. Powell, A., Shennan, S. & Thomas, M. G. Late Pleistocene demography and the appearance of modern human behavior. *Science* 324, 1298–1300 (2009).
46. Atkinson, Q. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332, 346–348 (2011).
47. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 6, doi:10.1101/pdb.top0408 (2010).
48. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-3.0
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 25, 1754–1760 (2009).
51. Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* 5, 337 (2012).
52. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).
53. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGS: analysis of next generation sequencing data. *BMC Bioinformatics* 15, 356 (2014).
54. Fumagalli, M., Vieira, F. G., Linderoth, T. & Nielsen, R. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* 30, 1486–1487 (2014).
55. Fumagalli, M. et al. Quantifying population genetic differentiation from next generation sequencing data. *Genetics* 195, 979–992 (2013).
56. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* 2, e190 (2006).
57. Fumagalli, M. Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS ONE* 8, e79667 (2013).
58. Reynolds, J., Weir, B. S. & Cockerham, C. C. Estimation of the coancestry coefficients for a SNP: a short Genetic distance. *Genetics* 105, 767–779 (1983).
59. Poelstra, J. W. et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344, 1410–1414 (2014).
60. Korneliussen, T. S. et al. Calculation of Tajima’s D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* 14, 289 (2013).
61. Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* 52, 863–867 (1987).
62. Ptak, S. & Przeworski, M. Evidence for population growth in humans is confounded by population structure. *Trends Genet.* 18, 559–563 (2002).
63. Gattepaille, L. M., Jakobsson, M. & Blum, M. G. B. Inferring population size changes from low depth next-generation sequencing data. *Genetics* 184, 523–533 (2009).
64. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57 (2008).
65. Zhan, S. et al. The genetics of monarch butterfly migration and warning colouration. *Nature* 514, 317–321 (2014).
66. Cavalli-Sforza, L. L. Human diversity. *Proc. 12th Int. Congr. Genet.* 2, 405–416 (1969).
67. Rebhan, M., Chaila-CSapi, V., Prilusky, J. & Lancer, D. GeneCard: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 14, 656–666 (1998).
68. Locasale, J. W. Serine, glycine and one-carbon unit: cancer metabolism in full circle. *Nat. Rev. Cancer* 13, 572–583 (2013).
69. Shao, Z. et al. Crystal structure of tRNA m1G9 methyltransferase Tm10: insight into the catalytic mechanism and recognition of tRNA substrate. *Nucleic Acids Res.* 42, 509–525 (2014).

Acknowledgements

The project was funded by a Marie Curie IF ‘KWAF10’ and Læsøki Foundation grants to A.D.F.; European Research Council grant ERC-2013-336336 to J.B.W.W.; a Danish National Research Foundation grant DNR994 to M.T.P.G.; by Swiss NSF Grant 31003A-144393 to L.E.; and by a short visit grant to A.D.F. from the ESF Research Networking Programme ConGenOmics. We acknowledge The Danish National High-Throughput DNA Sequencing Centre for sequencing the samples and, particularly, Andaine Seguin-Orlando, Lillian Petersen, Cecilie Demring Mortensen, Kim Magnussen and Ian Lissimore for technical support. Large-scale computational effort was made possible by the UPPMAX next-generation sequencing cluster and storage facility (UPPNEX), funded by the Knut and Alice Wallenberg Foundation and the Swedish National Infrastructure for Computing. Heng Li, Richard Durbin, Line Skotte, Anders Albrechtsen, Stephan Schüffers, Lounes Chikhi and Rasmus Nielsen provided useful feedback and discussions on the methods. We thank Willy Rodríguez for providing scripts for the ms simulations. Mike Ford and Kim Parsons provided useful feedback on an earlier draft of this manuscript. We thank Paul Fiedler for providing the map used in Fig. 1 and Uko Gorter for providing the illustrations of killer whale ecotypes, and Robert Pitman’s support and input throughout this project. We are grateful to M. Ford, K. Parsons, A. Burdin, M. Dahlheim, R. Pitman, the SWFSC Marine Mammal and Sea Turtle Research Collection, and to the International Whaling Commission for providing samples. We dedicate this work to the memory of Eva Saulitis, whose rich legacy includes an invaluable contribution to the foundations of our understanding of killer whale biology, without which this study would not have been possible.

Author contributions

A.D.F., P.A.M. and M.T.P.G. initially conceived and designed the study, which was further developed by J.B.W.W. and N.V., with input from all other authors; R.W.B., J.W.D., M.B.H. and P.W. conducted fieldwork and collected biopsy samples; K.M.R. and P.A.M. conducted DNA extraction and sample selection; A.D.F. conducted DNA library construction; J.W.D. and M.C.A.-A. conducted sequence read filtering and mapping; N.V., A.D.F., F.G.V., M.F., M.D.P., T.S.K., T.V. and V.C.S. ran the population genetic analyses; A.D.F. and N.V. wrote the manuscript and the Supplementary Information, with input from J.B.W.W., M.T.P.G. and the other authors.

Additional information

Accession codes: All sequencing data are available in European Nucleotide Archive (ENA) under accession numbers ER554424–ER554471 (see Supplementary Table 1 for a full list of accession numbers and associated sample ID).

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Foote, A. D. et al. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat. Commun.* 7:11693 doi: 10.1038/ncomms11693 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/