Science Advances
AAAS

Supplementary Materials for

Selection on ancestral genetic variation fuels repeated ecotype formation in bottlenose dolphins

Marie Louis*, Marco Galimberti, Frederick Archer, Simon Berrow, Andrew Brownlow, Ramon Fallon, Milaja Nykänen, Joanne O'Brien, Kelly M. Roberston, Patricia E. Rosel, Benoit Simon-Bouhet, Daniel Wegmann, Michael C. Fontaine, Andrew D. Foote, Oscar E. Gaggiotti

*Corresponding author. Email: marielouis17@hotmail.com

Published 27 October 2021, *Sci. Adv.* **7**, eabg1245 (2021)
DOI: 10.1126/sciadv.abg1245

This PDF file includes:

Supplementary text

Extended results

Genetic structure

The PCA (*38*) and NGSAdmix (*39*) analyses run on a set of 798,572 unlinked SNPs indicated that the three pelagic populations were more closely related to each other than the three coastal populations, which were much more differentiated from each other (Fig. 1, B and C, fig. S2 and S3). The NGSAdmix analysis showed a strong increase in logP between $K = 3$ and 4, which was also the greatest change in mean log likelihood (*75*) and almost no increase after $K = 6$ (fig. S2, A and B). With $K = 4$, all pelagic populations were grouped together while each coastal one represented a distinct cluster. With $K = 6$, each population represented a distinct cluster, but the eastern North Atlantic (ENA) pelagic population exhibited some admixture with the two other pelagic populations (fig. S2C).

The location of the samples within the PCA (Fig. 1C, fig. S3) (*38*) plot can be interpreted in terms of the mean pairwise coalescent time between each pair of samples (*86*), so that samples with greater covariance of alleles than the mean covariance among samples cluster together. Samples from the same population clustered together. These results suggest greater covariance of alleles among populations within the same ocean basin, and greater covariance of alleles among the pelagic populations than among coastal populations (Fig. 1C, fig. S3A, table S3). The patterns of differentiation suggest two possible evolutionary scenarios. They may indicate that coastal populations resulted from several founding events. However, they do not necessarily imply that they also originated from different source populations, shared variation between the two pelagic populations in the Atlantic indicated that the two Atlantic coastal populations were possibly derived from the same ancestral Atlantic population. On the other hand, the second axis supports a partially shared genetic basis of ecotype divergence among the three pairs. Therefore, another possible scenario is a single initial divergence into ancestral pelagic and coastal ecotypes followed by independent drift in each of the coastal populations and different levels of gene flow within each ecotype pair, as well as gene flow among pelagic populations. Based on our admixture plot results for $K = 2$ and 3 (fig. S2C) where the two Atlantic coastal populations show shared ancestry, this scenario appears plausible in the Atlantic.

Evolutionary relationships among ecotypes

TreeMix

Demographic history and potential admixture events were explored using TreeMix (*40*), in which shared co-variance of drift of allele frequencies suggests admixture events or gene flow (Fig. 2A, fig. S4 and S5). The maximum likelihood population trees indicated marginal drift in the pelagic populations from a shared ancestral population, as suggested by the short branch lengths, in contrast with the much longer branches and thus stronger drift in the coastal populations (Fig. 2A, fig. S5).

The best number of migration events was estimated to M=2, using an *ad hoc* analysis of the second order rate of change in the log-likelihood (Evanno) method (fig. S4). 99.8% of the variance was explained at M=2, and variance only increases marginally at M>2. This tree suggests a shared ancestral divergence between coastal and pelagic ecotypes with the WNAc and ENPc populations diverging first and a migration edge between the ENAc and WNAc populations suggesting shared variation (Fig. 2A). This divergence could then have been followed by more or less pronounced gene flow between coastal and pelagic ecotypes in each geographical region. Gene flow may be strong enough between the ENAc and ENAp populations to regroup them in the tree and between the ENPc and ENPp populations to be detected as a migration edge, but may be very low between ecotypes in the WNA allowing ancestral divergence to appear in the tree. The position of the Pacific populations remained relatively unresolved in this tree. The ENP coastal population diverged second after the WNA coastal population, but the subsequent divergence of the ENP pelagic population and Atlantic populations was only supported in 50-75% of the bootstraps. The alternative positions of those populations was the same as for M=0 and M=1 where the ENP pelagic and coastal ecotypes diverged from the same ancestral population. The uncertainty of the positions of the Pacific populations was further displayed in the trees with M>2, where both of the aforementioned evolutionary relationships are observed.

In this study, the covariance in allele frequencies between populations may be a consequence of low levels of on-going gene flow between populations at equilibrium rather than discrete migration events. These results and our other admixture test results reinforce the idea that highly mobile marine populations do not conform to a simple bi-furcating tree model (*57*). We therefore chose to apply mainly population genomic methods, which investigate variation in

allele frequencies and account for complex ancestry, rather than phylogenetic substitution-based inferences that assume a simple bifurcating branching process.

F_4 statistics

To further test for the best evolutionary model that fits a tree-like (bifurcation) process, we calculated the F_4 statistic, which can provide robust evidence of admixture, even if gene flow events occurred hundreds of generations ago and under scenarios of incomplete lineage sorting (*42*). Visualising the paths of genetic drift along the edges of a tree-like admixture graph can be an intuitive way to understand the expected value of the F_4 statistic (*42*). 'Drift' in this test is defined as the changes in allele frequencies along the graph edges, from population *a* to population *b (42)*. Therefore, the drift between *a* and *b* is a function of changes in allele frequencies between *a* and *b*. We refer to the path between *a* and *b* in the graph, as the drift path, following Patterson et al. (2012) (*42*). If we consider the equation F_4(a,b;c,d), if this is a true representation of a tree-like relationship among these four populations then there should be no overlap in the drift path from *a* to *b*, and the drift path from *c* to *d*. Thus, the difference in allele frequencies between populations *a* and *b* should be uncorrelated to the difference in allele frequencies between populations *c* and *d*, and F_4(a,b;c,d) = 0. We can thus test between the two competing hypotheses, F_4(pelagic$_x$, coastal$_x$; pelagic$_y$, coastal$_y$) and F_4(pelagic$_x$, pelagic$_y$; coastal$_x$, coastal$_y$) to see which is the best fitting model, or to identify if neither is a good fit for describing the relationships among these populations.

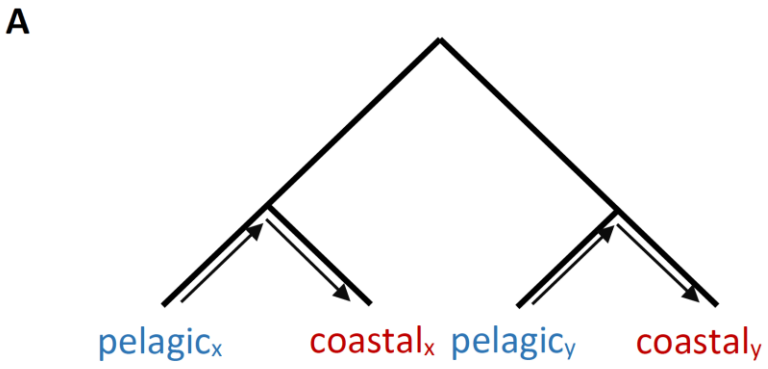

Fig. S1A. Tree of the form F_4(pelagic$_x$, coastal$_x$; pelagic$_y$, coastal$_y$).

Fig. S1A shows the paths along which drift is being measured, under the assumption of a tree-like relationship, in which F_4(pelagic$_x$, coastal$_x$; pelagic$_y$, coastal$_y$) is the true topology. If there is no overlap in the drift paths, F_4 = 0. Gene flow between pelagic$_x$ and coastal$_x$, or between pelagic$_y$ and coastal$_y$, would not affect this statistic. Only gene flow between geographic regions *x* and *y* would cause the statistic to deviate from zero. The F_4 statistic F_4(ENAp, ENAc;

ENPp, ENPc) is the closest to zero indicating the topology is a good, but not perfect representation of the relationships among these four populations.

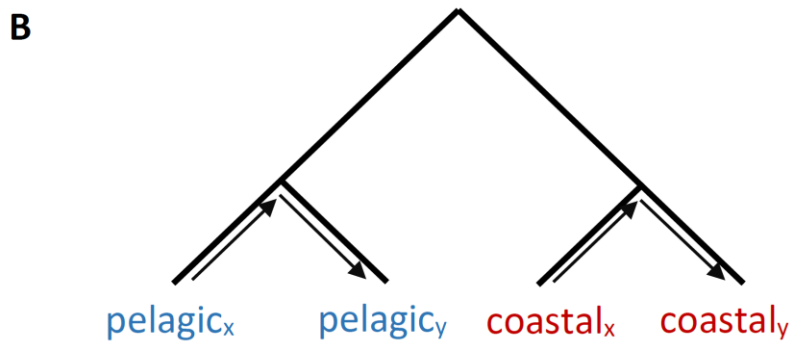

Fig. S1B. Tree of the form F_4(pelagic$_x$, pelagic$_y$; coastal$_x$, coastal$_y$)

Similarly, fig. S1B shows the paths along which drift is being measured, under the assumption of a tree-like relationship, in which F_4(pelagic$_x$, pelagic$_y$; coastal$_x$, coastal$_y$) is the true topology. There is no overlap in the drift paths and therefore $F_4 = 0$. Gene flow between pelagic$_x$ and pelagic$_y$, or between coastal$_x$ and coastal$_y$, would not affect this statistic. Only gene flow between coastal and pelagic ecotypes would cause the statistic to deviate from zero. The F_4 statistic F_4(ENAp, WNAp; ENAc, WNAc) is close to zero indicating the topology is a good, but not perfect representation of the relationship among these four populations.

If we now visualise some of the F_4 statistic results, comparing with the inference from TreeMix.

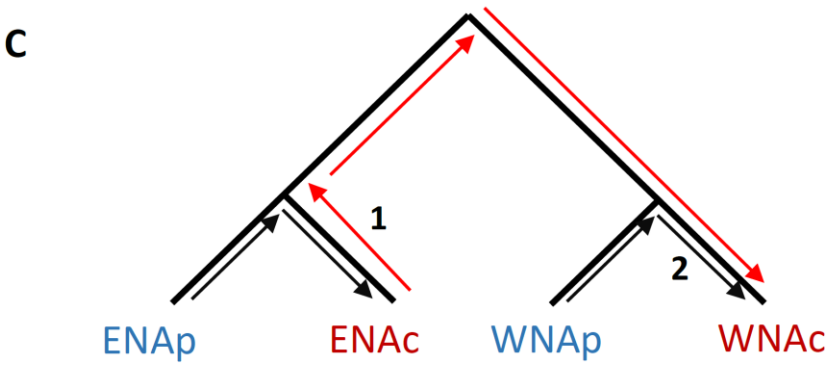

Fig. S1C. Tree of the form F_4(ENAp, ENAc; WNAp, WNAc), including drift paths and gene flow as estimated in TreeMix.

In fig. S1C we see the drift paths associated with the statistic F_4(ENAp, ENAc; WNAp, WNAc), represented by black arrows. Additionally, we visualise gene flow from the ENAc population to the WNAc population, as inferred by the best supported topology of TreeMix (Fig. 2A), as a red arrow. We see that gene flow between the ENAc and WNAc populations will result in some correlated changes in allele frequencies from ENAp to ENAc, and from WNAp to WNAc. Thus, we expect the F_4 statistic to deviate from zero. The direction of the gene flow along edge **1** is in the opposite direction to the drift along this edge and would reduce the F_4 statistic below 0. However, the direction of gene flow along edge **2** is in the same

direction as changes in allele frequencies due to drift and would increase the F_4 statistic F_4(ENAp, ENAc; WNAp, WNAc). This F_4 statistic would therefore be positive if the drift along edge 2 to WNAc was greater than the drift along edge 1 to ENAc, which is consistent with our results.

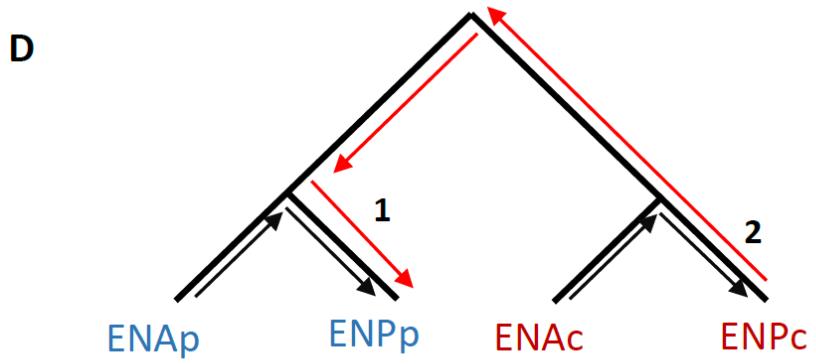

Fig. S1D. Tree of the form F_4(ENAp, ENPp; ENAc, ENPc), including drift paths and gene flow as estimated in TreeMix.

In fig. S1D we see the drift paths associated with the statistic F_4(ENAp, ENPp; ENAc, ENPc), represented by black arrows. Additionally, we visualise gene flow from the ENPc population to the ENPp population, as inferred by the best supported topology of TreeMix (Fig. 2A). We see that gene flow between the ENPc and ENPp populations will result in some correlated changes in allele frequencies from ENAp to ENPp, and from ENAc to ENPc. Thus, we expect the F_4 statistic to deviate from zero. The direction of the gene flow along edge **2** is in the opposite direction to the drift path along this edge, thus changes in allele frequencies caused by drift will be negatively correlated with changes in allele frequencies caused by gene flow. However, the direction of gene flow along edge **1** is in the same direction as changes in allele frequencies due to drift and would increase the F_4 statistic F_4(ENAp, ENPp; ENAc, ENPc). As this F_4 statistic is positive it would suggest shared drift between ENPc and ENPp prior to independent drift in each. Note that gene flow/shared history from ENAc into ENAp would also result in a positive F_4 statistic. A combination of these two may be driving a positive statistic for F_4(ENAp, ENPp; ENAc, ENPc).

Demographic history and patterns of ancestries

Ancient tracts analysis

Considering the genome as a mosaic of different evolutionary histories, we predict tracts of older ancestry components will have accumulated more mutations, which can segregate among populations. We hereafter refer to such tracts as ‘ancient’ given their older time to the most recent common ancestor (TMRCA). To identify putative ancient tracts in the dolphin genomes, we searched for clusters of dense private mutations (*53*) segregating in each coastal individual relative to the allopatric pelagic populations (see Material and Methods, fig. S11), taking variation in mutation rate along the genome into account.

The highest density of ancient tracts was found in the WNA coastal dolphins (tables S4 and S5), with the total length of all those tracts (20-25 Mb) at a posterior probability of $P > 0.8$, being higher than those found in the ENA and ENP coastal populations (10-15 Mb, table S4). Three to five times more inferred ancient tracts were shared between the coastal populations of the ENA and WNA than between the coastal populations of the ENA and ENP, and the WNA and ENP (table S6). This is in accordance with the partially shared ancestry of North Atlantic coastal populations inferred from both the F_4 statistics and TreeMix (Fig. 2, A and B). The time to most recent common ancestor (TMRCA) between the ancient tracts in coastal dolphins and the corresponding genomic regions in the outgroup had a Poisson distribution with a mean between 40 and > 100 thousand generations (~619,000-2,270,000 yBP, assuming a generation time of 21.1 years (*81*), table S5) depending upon the mutation rate assumed. The inferred TMRCA of these ancient tracts (0.6 to 2.3 million years) was much older than those of the rest of the genome (0.1 to 0.4 million years) (table S5). T_{Ancient} was the oldest for the WNAc ecotype (table S5). The divergence dates of *T. aduncus* and *T. truncatus* estimated by Moura et al. 2020 (*54*) and McGowen et al. 2020 (*55*) are close to the TMRCA of the ancient tracts found in the WNA coastal individuals (1.0-2.3 million years, table S5), after correcting for the different mutation rates used between studies. Evidence of ancestral introgression from extinct or unsampled (i.e. ghost) populations or species has recently been found in other marine species (e.g. killer whales and sea bass (*56*, 57)). However, these 'ancient' tracts do not need to be directly introgressed, but may rather have been retained as balanced polymorphisms (*17*). The presence of such ancient tracts may partially explain the basal position of the internal node of the branch leading to the WNA coastal population in TreeMix in our study (Fig. 2A, fig. S5) and in the *Tursiops truncatus* samples of a phylogenetic study using ~26,000 SNPs generated by RAD-sequencing (*54*).

Local PCA

The dolphin genomes were comprised of regions with different evolutionary histories, a pattern which was further corroborated using Local PCA (*50*). The three major evolutionary patterns inferred with Local PCA on PCs 1 and 2 (fig. S12) include: i) the pattern expected under the scenario of each coastal population originating from the pelagic population in the same region (corner 3, purple); ii) the pattern where the ENA and WNA coastal populations were more closely related than expected under independent ecotype splits on each side of the Atlantic (corner 1, green), as supported by the F_4-statistics and TreeMix (Fig. 2, A and B, fig. S5) and

iii) a pattern where the WNA coastal population was closer to the pelagic populations (corner 2, orange). Furthermore, on PCs 3 and 4, coastal populations from the Atlantic and Pacific clustered together and likewise for the pelagic populations, suggesting parallel ecotype-based processes (fig. S12).

Mechanisms of repeated evolution to coastal habitat

Flink

Overall, selective pressures were the most pronounced among coastal populations (Fig. 4A), although we acknowledge that divergent selection within coastal populations may be inflated by false positives associated with them having experienced strong drift, compared to the pelagic populations (Fig. 2A, Fig. 3). The coastal ecotype also exhibited the largest proportion of loci under homogenising selection. On the other hand, no loci under homogenising selection were detected between pelagic and coastal dolphins (CvsP value in Fig. 4A). The proportion of loci under homogenising selection in pelagic dolphins was significantly lower than that observed in coastal dolphins. In terms of divergent selection, its incidence in coastal populations was similar to that observed between pelagic and coastal populations, with the pelagic populations exhibiting the lowest incidence. As opposed to coastal dolphins, pelagic populations exhibited a similar proportion of neutral loci as that observed between coastal and pelagic ecotypes.

Plots of all neutral and selected raw genotypes show that pelagic populations tend to be homozygous for either the reference or alternated alleles as well as heterozygous, while the coastal populations are more homozygous for the reference allele, the reference genome being a WNA coastal individual (fig. S13B and S14B). This pattern suggests that variants inferred to be under selection in coastal populations were present as Standing Genetic Variation (SGV) in pelagic populations.

Almost all of the 89,796 SNPs under homogenising selection (97%) in the coastal populations were also polymorphic in the pelagic populations (86% in the ENAp, 85% in the ENPp, 82% in the WNAp and 68% in all three pelagic populations), supporting the hypothesis that selection has acted mostly upon standing genetic variation (SGV). Those 89,796 SNPs include closely linked SNPs (median distance of 69 bp) in 2,578 regions separated by at least 100 kb. Plots of

the 2D site frequency spectrum (SFS) of the SNPs under homogenising selection in the coastal populations also show that there was more shared variation between a pelagic and coastal population within a geographical region than when looking at all sites, i.e all SNPs in the genome (fig. S15). They also indicate more shared variation between the coastal populations for the SNPs under homogenising selection than for all sites (fig. S15), although there were still private alleles. Genotypes of the SNPs under homogenising selection in the coastal populations were more similar between ecotypes (fig. S16A) than those of all the SNPs (fig. S13B and 14B). In the PCA ran on the SNPs under homogenising selection in the coastal populations, the pelagic populations clustered together, and coastal populations were differentiated from them and from each other (fig. S16B). The first axis separated populations from the Pacific and the Atlantic. In the unrooted neighbor-joining tree (fig. S16C), generated from a distance matrix computed in the adegenet package in R (*87*), the two Atlantic populations clustered together, while the ENPc population clustered independently.

Most of the 89,663 SNPs under divergent selection (83%) between the ecotypes were also polymorphic in the pelagic populations (72% in the ENAp, 72% in the ENPp, 69% in the WNAp and 59% in all three pelagic populations), also supporting that selection has acted mostly upon SGV. Those 89,663 SNPs include closely linked SNPs (median distance of 108 bp) in 3,572 regions separated by at least 100 kb. Plots of the 2D-SFS under divergent selection between ecotypes show that there was less shared variation between a pelagic and a coastal population within a geographical region, than for all sites (fig. S15). In contrast, there was more shared variation among coastal populations for the SNPs under divergent selection between ecotypes than for all sites (fig. S15) although there were still a lot of private alleles. Genotypes of the SNPs under divergent selection between ecotypes were more dissimilar between the pelagic and coastal populations (fig. S17A) than those of all SNPs (fig. S13B and S14B). There was more fixed variation overall in the coastal populations, apart from the coastal population of the ENA, where there were many heterozygotes. The genotypes of the WNAc population stood out as mainly fixed for the reference genome (which is a WNA coastal individual). In the PCA ran on those SNPs, the pelagic populations clustered together with the ENPc, and the two coastal populations in the Atlantic were differentiated from them and from each other (fig. S17B). In the unrooted NJ tree, the dolphins clustered by ecotype, all coastal populations clustered together, and likewise for the pelagic populations (fig. S17C).

Most of the 7,165 SNPs under both homogenising selection among coastal populations and under divergent selection between the ecotypes (i.e. under parallel linked selection, 87%) were also polymorphic in the pelagic populations (72% in the ENAp, 76% in the ENPp, 69% in the WNAp and 57% were polymorphic in all three pelagic populations), again supporting selection acting mainly upon SVG. The plot of the genotypes shows relatively different patterns in the pelagic versus the coastal populations (Fig. 5A). Genotypes were more heterozygous in the coastal populations. In the PCA ran on those SNPs, the coastal populations were separated from the pelagic populations on the first axis with individuals from the same coastal population clustering together and apart from the other coastal populations on the second axis (Fig. 5B). In the unrooted NJ tree, the dolphins clustered by ecotype; all coastal populations clustered together, and so did the pelagic populations (Fig. 5C).

Thirty-three percent of the SNPs identified in the region as under homogenising selection were found in ancient tracts, while 61 and 66% of the SNPs under divergent selection between ecotypes, and under both types of selection were respectively found in ancient tracts (fig. S19). In contrast, only an average of 20%, 20% and 22% of putatively neutral SNPs were found in ancient tracts in 100 random samples of the same number of SNPs in homogenising, divergent and parallel linked selection, respectively (fig. S19).

We have also plotted the 1D-SFS for the 7,165 SNPs under parallel linked selection (fig. S18), and the variable sites show an excess of intermediate variants for the coastal populations, and have a shape which is clearly different from those of all SNPs (fig. S9). In contrast, the variable sites of the 1D-SFS of the three pelagic populations show an excess of rare variants for the 7,165 SNPs as also observed for all SNPs.

The 7,165 SNPs putatively under parallel linked selection overlapped with 45 genes. Those include genes related to cognitive abilities, learning and memory (RELN (*61*, *62*), ADER3 (*63*)) neuronal activity regulation (INSYN2A), lipid metabolism (AGK, LPIN2 (*65*), KLB), muscle contraction (RYR1, myosin-3, myosin-13, CAMK2D), axe growth (FEZ1, FEZ2), heart functions (CAMK2D), tooth enamel development (MMP20 (*88*)), immunity (HLA class II histocompatibility antigen, DQ alpha 2 chain, SERINC5), oxidative stress (cytochrome b5 reductase 4) and hormone regulation (STAR, table S7).

Extended materials and methods

Sample collection

Epidermal tissue samples were collected from 57 bottlenose dolphins (Fig. 1A, table S1). Samples from the ENA were stranded animals found dead on the beach and assigned to the pelagic and coastal ecotypes in a previous study using microsatellites and a portion of the mitochondrial control region data (*26*). Coastal dolphins included five previously photo-identified bottlenose dolphins on the east coast of Scotland. Samples from the WNA and the ENP were collected as per Rosel *et al.* 2009 (*34*) and Lowther-Thieleking *et al.* 2015 (*28*) respectively. Samples were previously assigned to either coastal or pelagic ecotypes (*26*). Samples were stored at −80°C with no preservative or at −20°C fixed in a salt-saturated 20% DMSO solution or ethanol or with no preservative.

Laboratory procedures

DNA was extracted from epidermal tissue using a Qiagen DNeasy kit following the manufacturer's protocol for the ENA samples. For the WNA samples, DNA was extracted using standard proteinase K digestion and organic extraction as described in Rosel and Block 1996 (*89*). For the ENP samples, DNA was extracted using a sodium chloride protein precipitation (*90*).

Illumina libraries were built on 300-bp DNA fragments and were pooled equimolarly by geographical regions. The geographic region pools were sequenced across six lanes of an Illumina HiSeq X Ten platform for the ENA and ENP (20 DNA samples each) and five lanes for the WNA (17 DNA samples) using paired-end 150-bp chemistry. Leakage of reads between indexes in pooled samples on the same lane of Illumina sequencing platforms has been reported in the literature (*91*). We are confident that leakage is not an issue within our dataset as we found almost no ambiguities in our unpublished mitogenome dataset, not presented here (12 sites where a single nucleotide did not represent 100% of the reads, out of 934,287 sites, i.e. 16,391 bp for 57 individuals, that is 0.001%). In addition, we are confident the data included in this study do not present biases linked to data generation as they were generated at the same time, using the same library build protocol and the same sequencing platform.

Data processing and filtering

Read trimming and mapping

Demultiplexing was performed by BGI. Sequencing reads were processed with Trimmomatic v. 0.32 (*67*) to trim residual adapter sequence contamination using the default options for the seed mismatches: 2, the palindrome clip threshold: 30 and the simple clip threshold: 15 and low quality bases. We removed low quality bases (a phred score of less than 5) from the beginning and the end of the reads. In addition, we also performed a sliding-window trimming, cutting bases when the average quality within a window of 4 bp fell below a phred score of 15 (default parameter). Sequence reads that were less than 75 bp were discarded.

The remaining filtered reads were first mapped to a bottlenose dolphin mitochondrial genome (Genbank gi_557468684_gb_KF570351.1_.fasta) (*68*) as per Morin et al. 2015 (*92*). Reads that did not map to the mitochondrial genome were then extracted from the bam files and converted into a fastq file using samtools v. 1.2 (*37*, *76*) and picard-tools v. 2.1.0 (*70*). These reads were then mapped to the reference bottlenose dolphin genome assembly (Genbank: GCA_001922835.1, NIST Tur_tru v1) using BWA mem (v. 0.7.15) with default options *(69*).

Data filtering

We checked and confirmed the quality of our data using FASTQC (*93*) after trimming, mapping and filtering. Picard-tools v. 2.1.0 (*70*) was used to add read groups and merge the bam files from each individual from the different lanes and remove duplicate reads. The optical duplicate pixel distance was set to 2,500 as recommended by the Broad Institute to better estimate library complexity on data generated using the Illumina HiSeq XTen platform. Then, indel realignment was performed using GATK v. 3.6.0 (*71*). Samtools was used to keep only the mapped reads with a mapping quality of at least 30.

Repeat regions from the cetartiodactyla group were identified in the common bottlenose dolphin reference genome using RepeatMasker (*72*) and saved in a bed file. Only interspersed repeats were masked and STRs, small RNAs and low complexity regions were retained. The repeat regions were removed from the bam files using bedtools v. 2.25.0 (*94*) and samtools v. 1.2. We also removed regions of excessive coverage as the high coverage of these regions can potentially be the result of unmasked repeated regions, in nuclear mitochondrial DNA (NUMTs), or some other mapping artifact (e.g. paralogous loci). Coverage was then estimated for each genome using the doDepth function in ANGSD v. 0.913 (*73*). However, due to the size of the dataset, global coverage was estimated by randomly sampling three individuals per

ecotype (i.e. for a total of 18 individuals). Regions that were higher than twice the mean coverage (>346x) were considered of excessive coverage. These regions were detected using the CALLABLELOCI tool in GATK. Then they were removed from the bam files using bedtools and samtools as above. Mean coverage was again estimated using the doDepth function in ANGSD.

To identify the scaffolds corresponding to the autosomes and the X chromosome, we randomly sampled five males and five females and estimated mean coverage for males and females for all scaffolds that were longer than 1 Mb (i.e. 98.3% of the genome). When the ratio of the mean coverage for the females on the mean coverage for the males was around 1 (mean=1.06, min=1.04, max=1.19), it was considered the scaffold corresponded to an autosome and when the ratio was around 2 (mean=2.05, min=1.96, max=2.07), the scaffold was considered as belonging to the X chromosome. Twelve scaffolds > 1 Mb were identified as corresponding to the X chromosome and 116 scaffolds to the autosomes. The 12 scaffolds belonging to the X-chromosome were removed using bedtools and samtools. When the sex was unknown, it was identified by estimating coverage for the 12 X-chromosome scaffolds and 12 scaffolds from the autosomes.

ANGSD was used to identify SNPs that show significant deviation from HWE and a F value <0 within populations as they can be the result of paralogs or other mapping artefacts. The latter sites were removed from the bam files using bedtools and samtools as above.

Linkage Disequilibrium (LD) pruning

NgsLD (*74*) was used to obtain a set of unlinked SNPs. It was first run with a maximum distance of 1,000 kb for SNPs to be possibly in LD, randomly subsampling 5% of the data and LD decay was inspected using R. As LD decay was decreasing rapidly, NgsLD was then re-run using a maximum distance of 100 kb. LD decay was inspected and showed LD was negligible after a distance of 20 kb. Then, a set of unlinked sites was produced considering that SNPs are in LD until 20 kb and using a minimum weight of 0.5. Population structure analyses were run on the set of unlinked SNPs. All the other analyses (demographic history or selection) were run on the full sets of SNPs.

Ancestral state reconstruction

The ancestral state of the alleles was reconstructed by creating a consensus sequence using two whole genomes of the killer whale (*Orcinus orca*), the sperm whale (*Physeter macrocephalus*) and the finless porpoise (*Neophocaena phocaenoides*). Short read data from two killer whales, a sperm whale and a finless porpoise (SRR574982/SRX188934, SRR1162264/SRX447351, SRR1031998/SRX378812, SRR940959/SRX326372 (*83*, *95*–*97*)) were additionally accessed from the National Center for Biotechnology Information Sequence Read Archive database and mapped to the common bottlenose dolphin reference genome assembly as described above for the modern bottlenose dolphin samples. Slight changes included: i) setting the phred score to 20 instead of 15 for the sliding window in Trimmomatic, as base quality of the raw data were lower than for the bottlenose dolphin data and ii) setting the minimum mapping quality to 20 as mapping in against another species. Coverage was estimated for each genome using the doDepth function in ANGSD and they were then subsampled to a coverage of 5x using samtools. The four genomes were merged using samtools and then the consensus sequence was inferred by selecting the most common base using the doFasta 2 option together with doCounts 1 in ANGSD. Sites not present in the ancestral state but present in our dolphin re-sequencing data were masked for the estimation of the site frequency spectrum (SFS) and the 2D-SFS.

Admixture analyses

We reconstructed the relationships among coastal and pelagic bottlenose dolphins in the ENA, WNA and ENP using admixture and 'treeness' tests. TreeMix and F_4 statistics explicitly test for admixture and can also inform on the directionality of gene flow (*42*).

TreeMix estimates a bifurcating ML tree based on genome-wide population allele frequency data and uses a Gaussian approximation to estimate genetic drift among populations. The relationships among populations are represented by the branches of the tree based on the majority of alleles. Migration edges are fitted between populations that are a poor fit to the tree model from the covariance matrix of allele frequencies. The direction of the migration events is inferred from asymmetries in the covariance matrix of allele frequencies relative to an ancestral population.

F_4 statistics quantifies drift (i.e. changes of allele frequencies) between pairs of populations in a tree (*41*, *42*). The relationships between four populations can be described by three possible unrooted trees. For example, the relationships between populations A, B, C, D could be

represented by three trees (A,B;C,D), (A,C;B,D) and (A,D;B,C). When the topology is correct, the difference in allele frequencies (i.e. the drift that has accumulated) between the two populations in each clade should be uncorrelated between clades. The F_4 statistics F_4(A,B;C,D) would thus not differ significantly from 0. For incorrect topologies, correlated drift would lead to significantly positive or negative correlation values. To test whether each geographic pair of pelagic and coastal ecotypes had evolved independently, we estimated F_4(pelagic$_x$, coastal$_x$; pelagic$_y$, coastal$_y$), and to test for a shared colonisation history of both the coastal and pelagic ecotype within a geographic region we estimated F_4(pelagic$_x$, pelagic$_y$; coastal$_x$, coastal$_y$).

Demographic history

SMC++

Demographic history, that is changes in effective population sizes (N_e) through time and ecotype splits within a region, were computed using the program SMC++ v. 1 (*46*), i.e. Sequential Markov Coalescent + plenty of unlabelled samples. The method incorporates both the site frequency spectrum (SFS) and Linkage Disequilibrium information in a coalescent Hidden Markov Model, HMM (similar to PSMC, Pairwise Sequentially Markovian Coalescent (*98*)). SMC++ is an extension of the PSMC for a larger number of samples. While PSMC uses the distribution of heterozygous sites throughout the genome where the heterozygosity information is emitted as binary, SMC++ emits the allele frequency of an extra n-2 haplotypes. The latter is based on the SFS conditioned on the TMRCA of a single 'distinguished individual'. It can include several individuals per population while PSMC analyses are only based on one individual and phasing the data is not required.

Then, only the autosome scaffolds, which were more than 10 Mb were included in the analyses, and no MAF filter was applied on the vcf file. The vcf file was converted to SMC++ format using the vcf2smc function for each retained scaffold. The repeated regions and excessive coverage regions were included as a mask file so that they were not misidentified as very long runs of homozygosity which could impact the population trajectories and create false sudden decreases in N_e in recent times. The analysis was run both using all regions and taking out all the regions under selection, as identified with Flink (see below). Regions under selection were defined as 50 kb around each outlier SNPs (thus 25 kb each side). Regions under selection were included in the mask file when they were taken out from the dataset.

As mentioned earlier, SMC++ estimated the SFS conditioned on the TMRCA of a single individual, called “distinguished” individual hereafter. We fixed the distinguished individual to i) a particular individual, or ii) made it vary over two individuals and iii) made it vary over three individuals. Varying the distinguished individual over different individuals has the advantage of incorporating genealogical information from additional individuals into the analysis, which may lead to improved estimates.

Population size histories were estimated using the *estimate* option in SMC++ using the default settings for the estimate function, a generation time of 21.1 years for the species (*81*) and two different mutation rates. Mutation rates were i) 9.10e-10 substitutions per site per year that is 1.92e-8 substitution per nucleotide per generation (*82*) and ii) 1.21e-9 substitution rate per site per year (*83*) that is 2.56e-8 substitution per nucleotide per generation.

Population split estimations first involved estimating population histories using the estimate option, i.e. the marginal estimates. Then, datasets containing the joint frequency spectrum for both populations were computed using the vcf2smc function. Lastly, the *split* function was used to refine the marginal estimates into an estimate of the joint demography and divergence times were estimated between ecotypes in each region. Results were plotted in R v. 3.6.1 (*84*) with packages ggplot2 (*99*), scales (*100*) and RcolorBrewer (*101*)).

Diversity statistics

Nucleotide diversity, Theta Watterson and Tajima’s D were estimated for each population using ANGSD v. 0.921 (*73*, *102*). First, the unfolded site frequency spectrum (SFS) was computed for each population in ANGSD using a two step procedure (*102*) for sites with data in all individuals and including the ancestral state as defined earlier. First, the dosaf 1 function was used to calculate the site allele frequency spectrum likelihood (saf) based on individual genotype likelihoods assuming HWE. Then, the realSFS function was used to optimize the saf and estimate the SFS. Nucleotide diversity and Theta Watterson were calculated for each site and then both the latter and Tajima’s D were estimated from the SFS using a sliding-window size of 50 kb and a step size of 10 kb. To compute the 2D-SFS, the realSFS function was run on the saf files from each pair of populations.

Selection analyses

In Flink (*47*), we used the values used to analyse the human data in the method's paper as initial values for the choice of the range for each parameter. The function estimate was run with parameters A_max (maximum coefficient of selection) set to 4.0, B (coefficient of drift of a group) to a mean of -2.0 and standard deviation of 1.8, lnK (logarithm of positive scaling parameter) of a minimum of -10.0 and maximum of -0.1, lnMu (Probability involved in the generating matrix to go to a different state for the higher hierarchy) to -4.0, 0.0; lnNu (Probability involved in the generating matrix to go to a selection state from the neutral state for the higher hierar) to -5.0, 0.0 - apart for scaffold groups 22 and 23 where it was reduced to -4.0 due to convergence issues, lnMu_g (Probability involved in the generating matrix to go to a different state at the group level) to -4.0, -0.0, lnNu_g (Probability involved in the generating matrix to go to a selection state from the neutral state at the group level) to -5.0,-0.0, s_max: maximum state of the Markov model to 14, beta (coefficient of drift of a population) to a mean of -2.0 and standard deviation of 1.8, alpha_max (maximum coefficient of selection of a group) = 4.0, lnKappa (logarithm of group positive scaling parameter) of a minimum of -10.0 and maximum of -0.1, sigmaProp_mu (Value to determine the range of the proposal value of mu) to 0.005, sigmaProp_nu (Value to determine the range of the proposal value of mu) to 0.05, sigmaProp_kappa (Value to determine the range of the proposal value of lnkappa) to 0.05. The number of iterations was set to 500,000, the burn-in to 300,000 and the thinning to 100. Convergence was verified in the results for all the parameters.

To get further insights into the results obtained by Flink, we plotted the raw genotypes of all the SNPs, SNPs under homogenising selection in the coastal populations, SNPs under divergent selection between ecotypes, and SNPs under both homogenising selection in the coastal populations and divergent between ecotypes (defined as the SNPs under parallel linked selection) using the R packages vcfR (*103*) and adegenet (*87*). We also plotted a neighbor-joining distance tree for the SNPs under each type of selection using the R package ape (*104*) and a PCA using the packages adegenet (glPCA function) and scales (*100*). To determine the origin of the SNPs under selection, we defined how many were also polymorphic in the pelagic populations, and compared the 2D-SFS between all pairs of populations, estimated in ANGSD, using all SNPs, the SNPs under homogenising selection in the coastal populations, SNPs under divergent selection between ecotypes. We plotted the 1D-SFS for the SNPs under parallel linked selection. Then, we defined how many SNPs under the two types of selection and under

parallel linked selection were found in ancient tracts. We compared the results with 100 random samples of the same number of putatively neutral SNPs found in ancient tracts.

Putative functions of the genes overlapping with the SNPs putatively identified as under parallel linked selection in Flink, were determined using literature search, Entrez Gene (*105*), Uniprot (*106*), RefSeq (*107*), GeneCards(*108*) and Online Mendelian Inheritance in Man (OMIM)(*109*) databases.

Supplementary figures

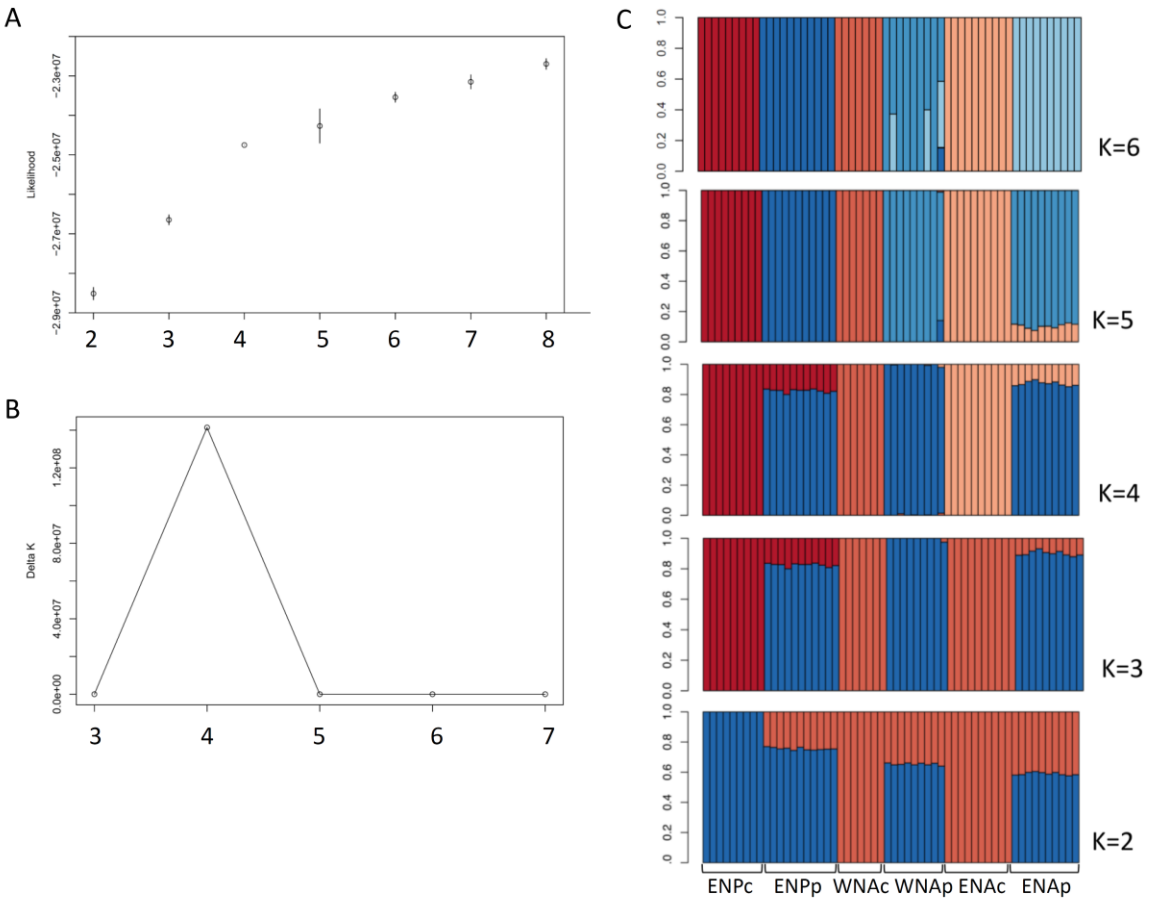

Fig. S2. (A) NGSAdmix log likelihoods and (B) the rate of likelihood change (Delta *K*) (*75*) for each number of cluster (K) values. (C) Ancestry proportions for each of the 57 individuals inferred in NGSAdmix for K=2 to *K*=6.

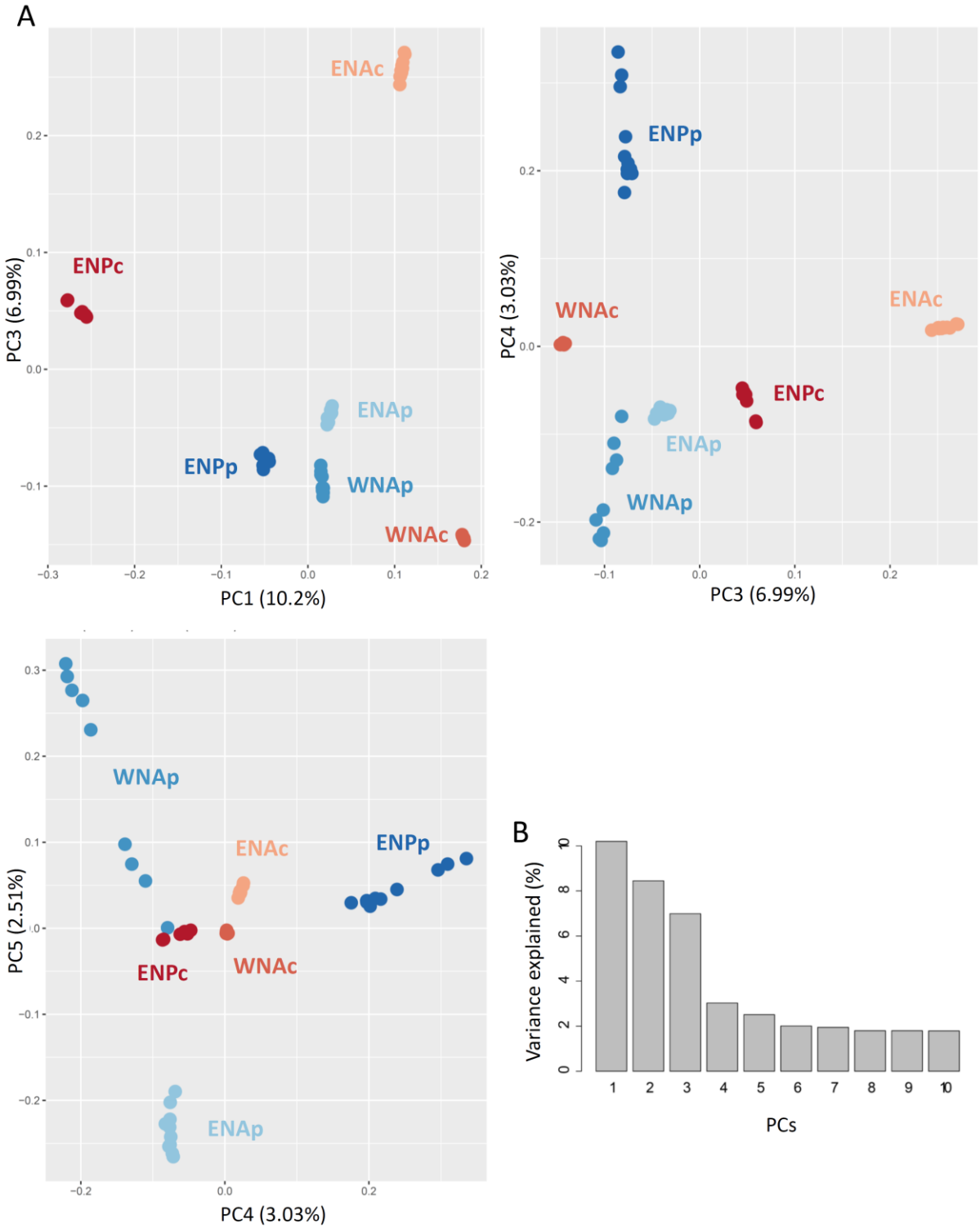

Fig. S3. (A) Principal component analysis (PCA) showing first and third, third and fourth, and fourth and fifth principal components (PCs). (B) PCA scree plot showing the percentage of variance explained by each of the 10 first principal components (PCs).

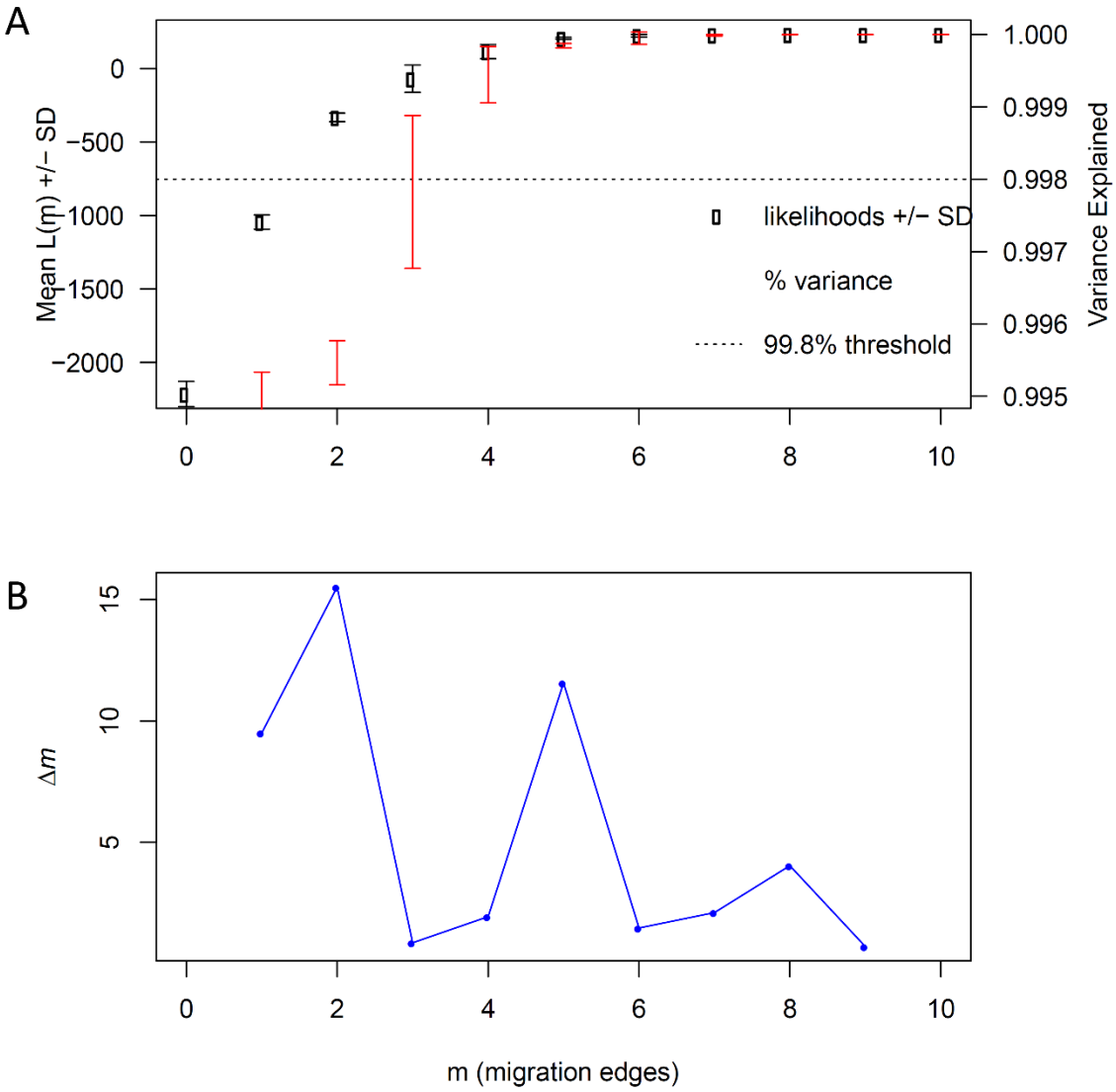

Fig. S4. Determination of the optimal number of migration edges between 0 and 10 in TreeMix using the (A) log-likelihood values and percentage of variance explained, and (B) the second order rate of change in the log-likelihood (Evanno) method.

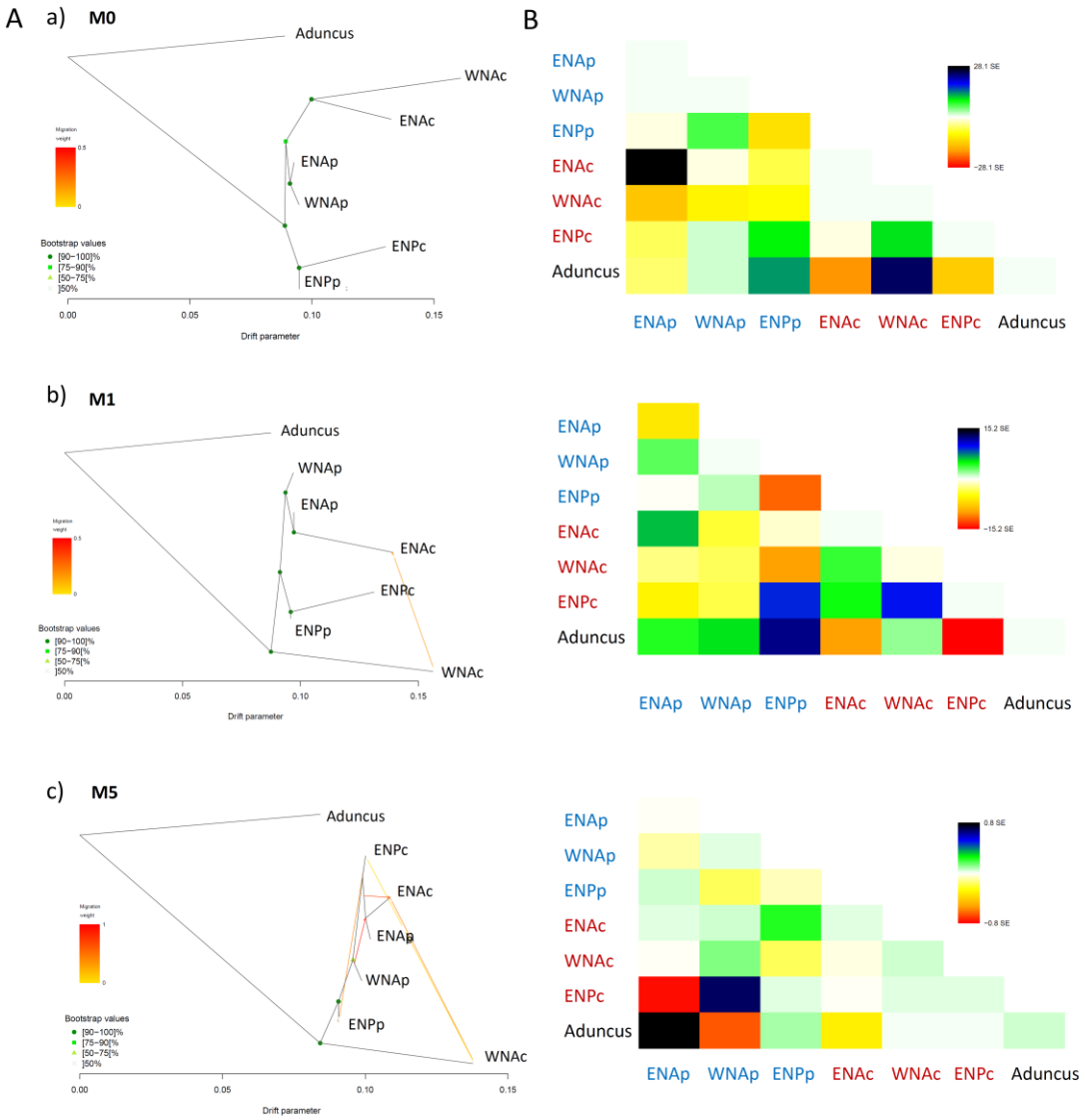

Fig. S5. (A) TreeMix consensus tree and bootstrap values displaying the relationships among populations as a bifurcating maximum-likelihood tree for a) no migration edge (M0), b) one migration edge (M1) and c) five migration edges (M5). Horizontal branch lengths represent the amount of genetic drift that has occurred along each branch. (B) Residual fit of the observed versus the predicted squared allele frequency difference, expressed as the number of SE of the deviation. SE values are represented by colours according to the palette on the right. Residuals above zero indicate populations that are more closely related to each other in the data than in the best-fit tree and have potentially undergone admixture. Negative residuals represent populations that are less closely related in the data than represented in the best-fit tree.

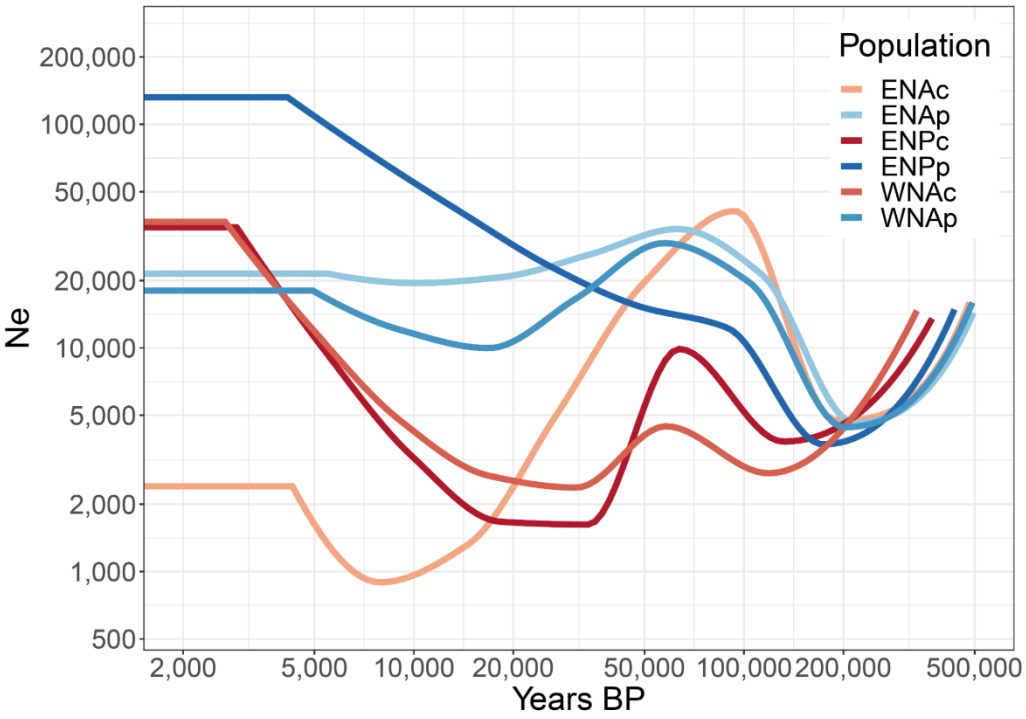

Fig. S6. Changes in effective population size through time inferred for each bottlenose dolphin population using SMC++ using a mutation rate of 1.92e-8 substitution per nucleotide per generation (*82*) and a generation time of 21.1 years (*81*). Only neutral sites are included.

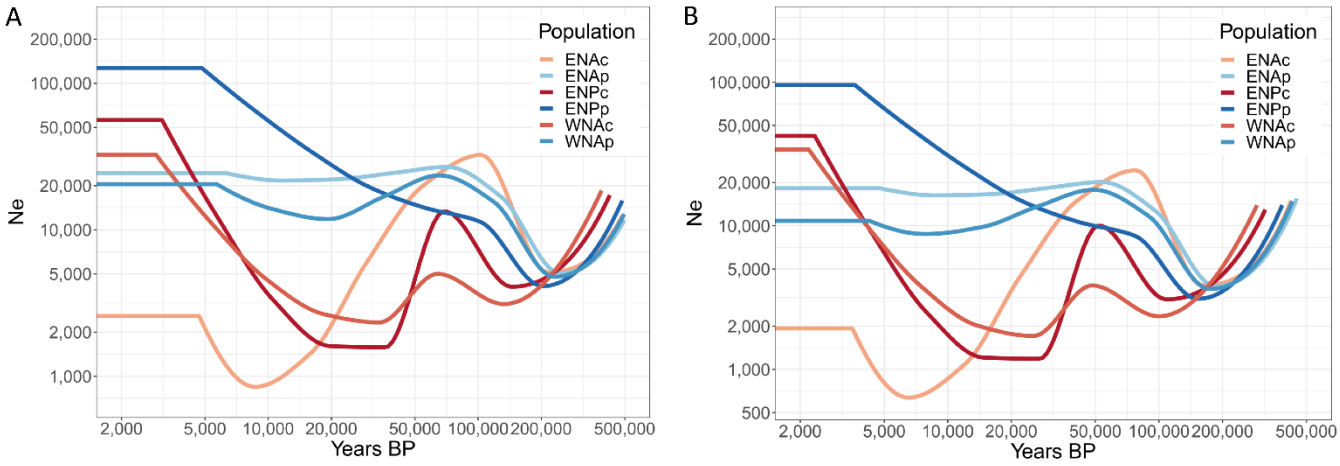

Fig. S7. Changes in effective population size through time inferred for each bottlenose dolphin population for all sites (neutral and under selection as identified by Flink) using SMC++, a generation time of 21.1 years (*81*), and (A) a mutation rate of 1.92e-8 substitution per nucleotide per generation (*82*) and (B) a mutation rate of 2.56e-8 substitution per nucleotide per generation (*83*).

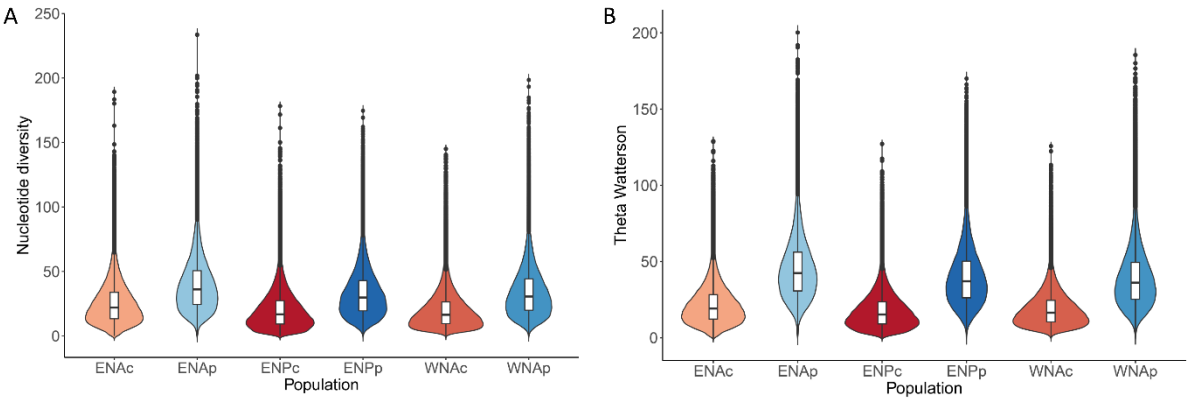

Fig. S8. (A) Nucleotide diversity and (B) Theta Watterson estimated for each population of bottlenose dolphins from the site frequency spectrum (SFS).

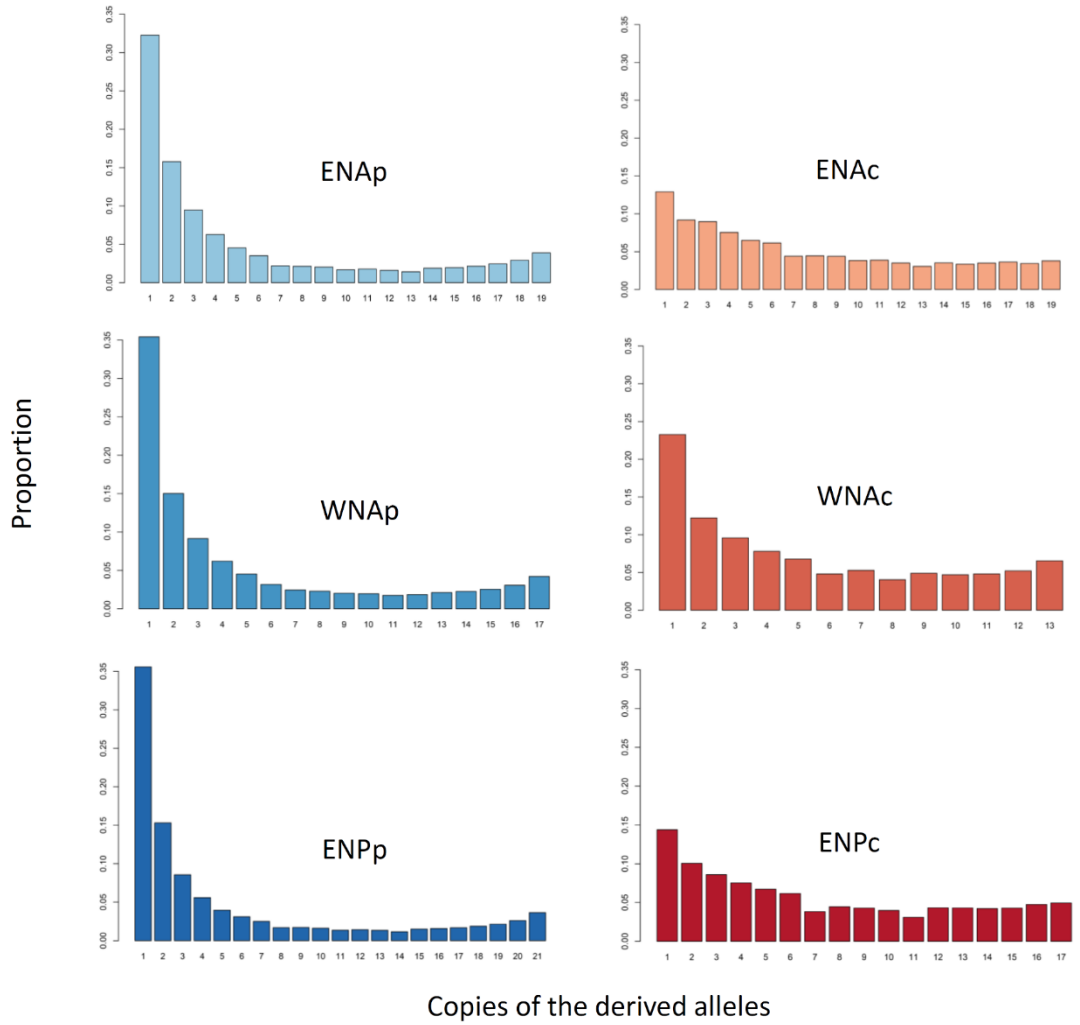

Fig. S9. Site frequency spectrum (SFS) for each of the bottlenose dolphin population, as the proportion of each number of copies of the derived alleles.

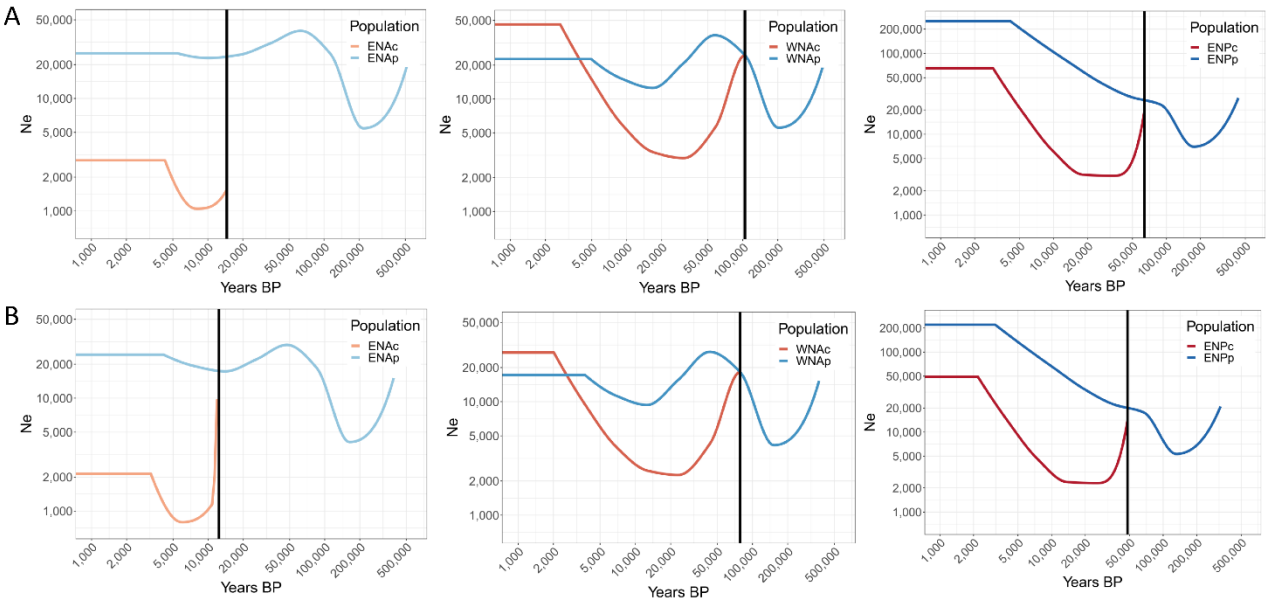

Fig. S10. Divergence time between ecotype pairs in the ENA, WNA and ENP estimated using SMC++, a generation time of 21.1 years (*81*), and (A) a mutation rate of 1.92e-8 substitution per nucleotide per generation (*82*) and (B) a mutation rate of 2.56e-8 substitution per nucleotide per generation (*83*).

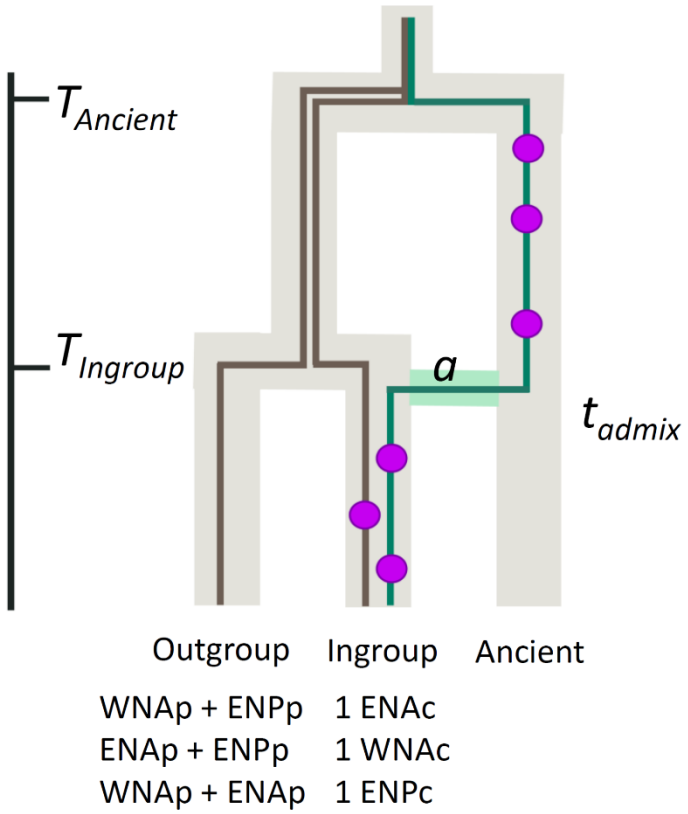

Fig. S11. Ancient ancestry method's principles, re-drawn from Skov et al. 2018 (*53*). At time T_{admix}, an ancient tract introgressed from a ghost population into the ingroup population, here one of the coastal population, with admixture proportion *a*. We test for ancient introgressions into each coastal population. All non-allopatric pelagic samples are part of the outgroup. The method scans the genome for clusters of private alleles (purple circles) in each coastal population. Ancient introgressed regions will have higher private variants density than non-introgressed. This is because the split between the pelagic populations and the ghost population – $T_{Ancient}$ – is older than the split between the pelagic and coastal populations $T_{Ingroup}$. Therefore, ancient tracts have had more time to accumulate variation not found in the pelagic populations. However, we hypothesise that ancient tracts in coastal bottlenose dolphin populations may not have been directly introgressed but rather have been retained as balanced polymorphism, see results and discussion.

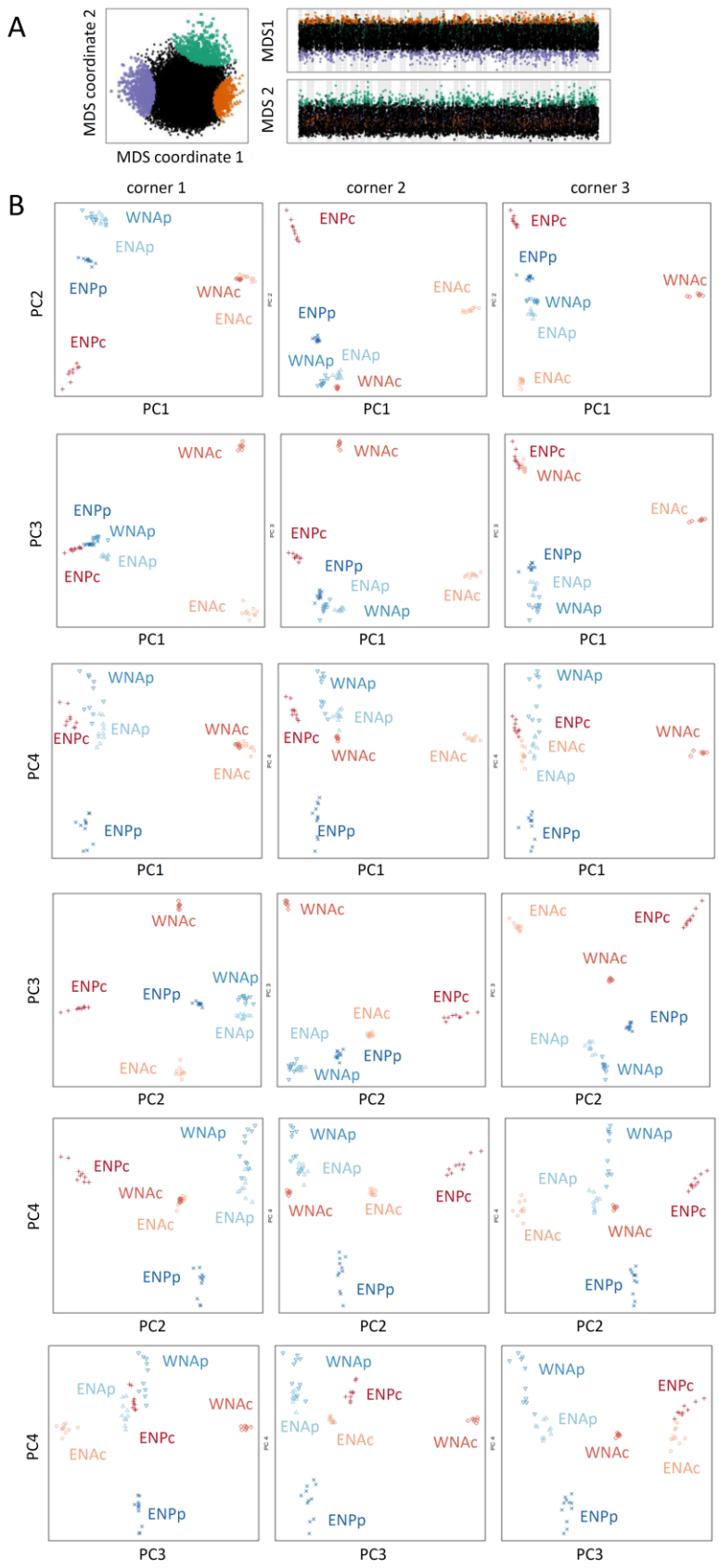

Fig. S12. Local PCA results. (A) MDS identifying the three major patterns of relatedness among bottlenose dolphin populations and (B) PCA describing the three major patterns of relatedness (corners 1-3, green, orange and purple respectively) among populations on four PCs on the 56 scaffolds > 10Mb.

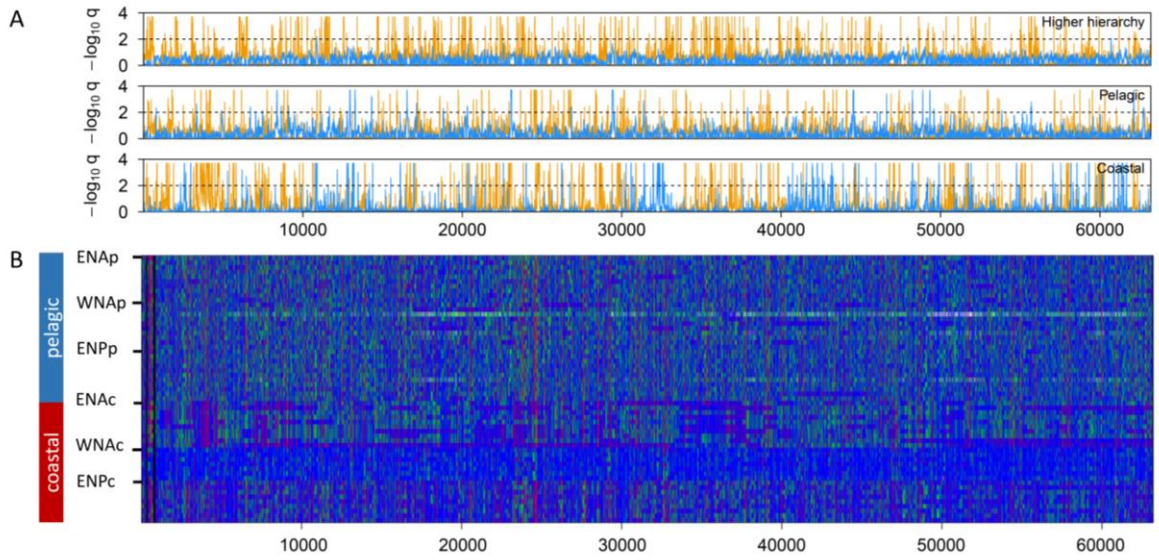

Fig. S13 (A) Patterns of selection (divergent: yellow, homogenising: blue) inferred using Flink for the different hierarchical groupings which are between pelagic and coastal populations (top panel), among pelagic populations (middle panel) and among coastal populations (lower panel) for super-scaffold 4. The y-axis indicates the locus-specific FDR for divergent (orange) and homogenising (blue) selection, respectively. The black dashed line shows the 1% FDR threshold, above which we consider the locus under selection. (B) Plot of the genotypes along scaffold ensemble 4, with blue: homozygote reference, green: heterozygote, and red: homozygote alternated. Black lines indicate the separation between different scaffolds.

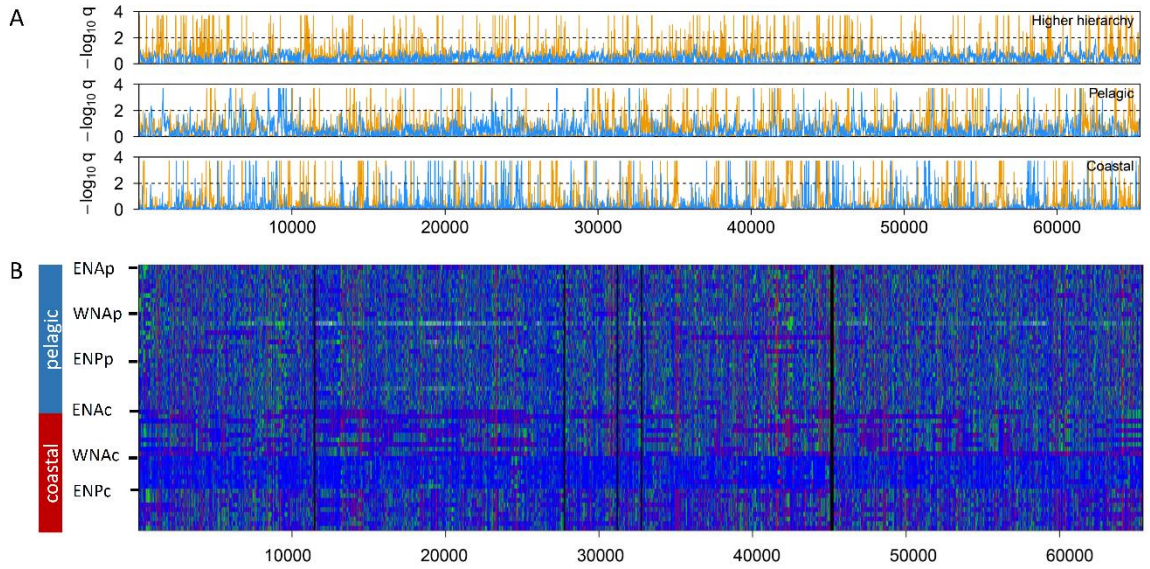

Fig. S14 (A) Patterns of selection (divergent: yellow, homogenising: blue) inferred using Flink for the different hierarchical groupings which are between pelagic and coastal populations (top panel), among pelagic populations (middle panel) and among coastal populations (lower panel) for super-scaffold 9. The y-axis indicates the locus-specific FDR for divergent (orange) and homogenising (blue) selection, respectively. The black dashed line shows the 1% FDR threshold, above which we consider the locus under selection. (B) Plot of the genotypes along scaffold ensemble 9, with blue: homozygote reference, green: heterozygote, and red: homozygote alternated. Black lines indicate the separation between different scaffolds.

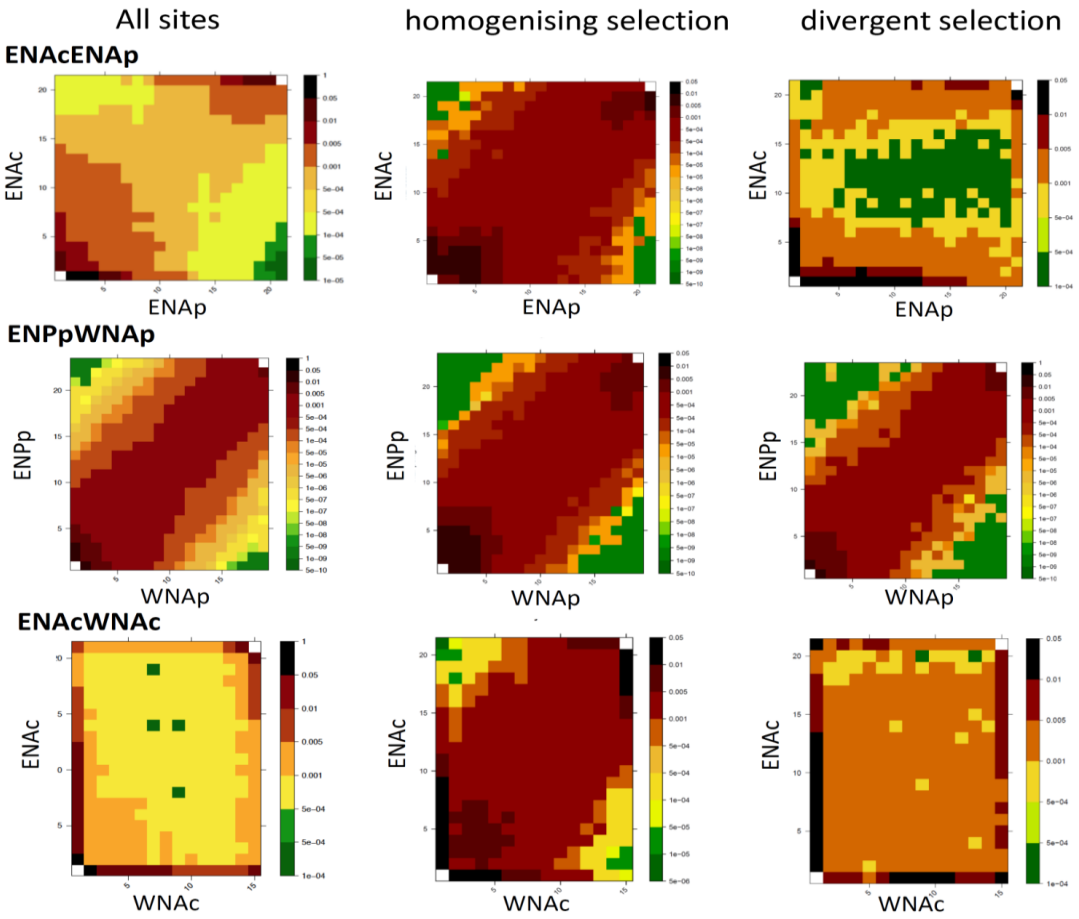

Fig. S15. 2D-site frequency spectrum (2D-SFS) for all sites (left panels), sites under homogenising selection in the coastal populations (middle panels), sites under divergent selection between ecotypes (right panels) for a coastal and a pelagic population (ENAc-ENAp - first row), two pelagic populations (ENPp-WNAp - middle row) and for two coastal populations (ENAc-WNAc - bottom row). Note that the patterns seen in these populations are similar across all pairs. Note that the scale differs for each plot due to large differences in the values included.

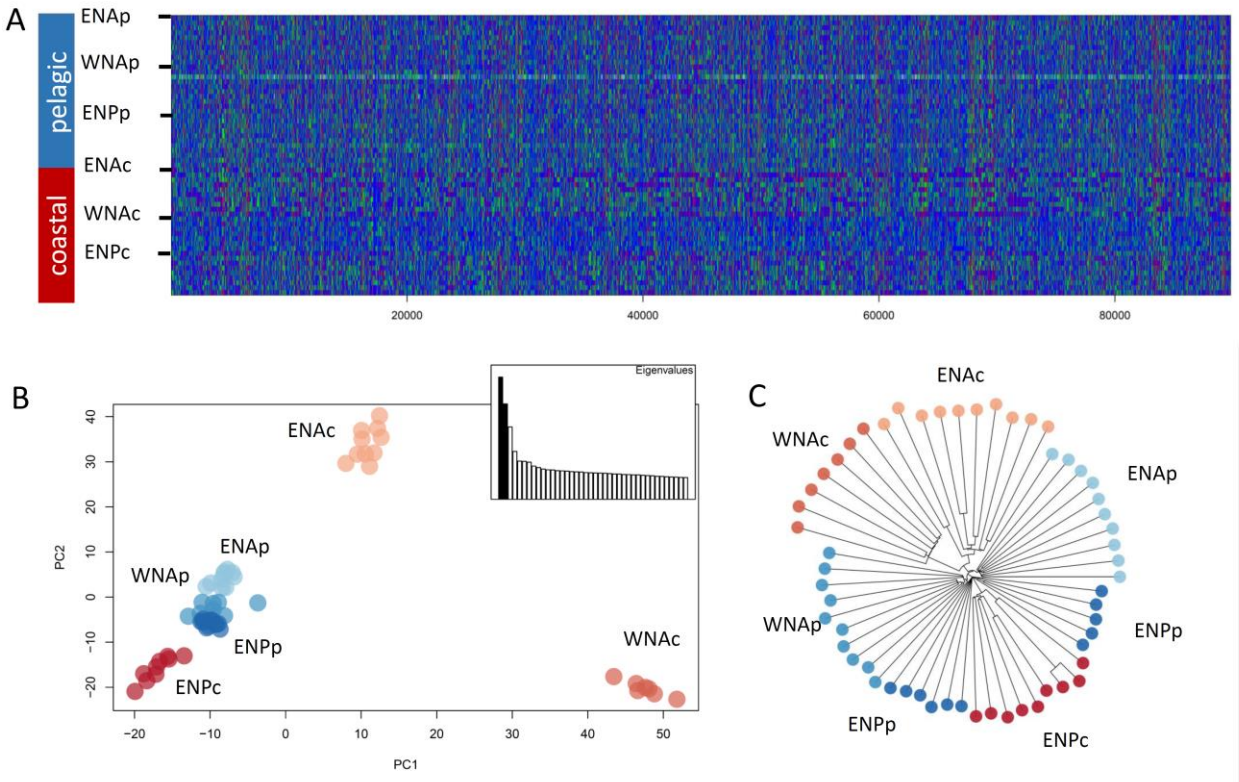

Fig. S16. Patterns of genetic variation of the 89,796 SNPs, under homogenising selection among the coastal populations, including closely linked SNPs scattered across the genome in 2,578 regions separated by at least 100 kb. (A) Plot of the genotypes, with blue: homozygote reference, green: heterozygote, and red: homozygote derived, (B) Principal component analysis and (C) Neighbor-joining distance tree of the common bottlenose dolphin samples.

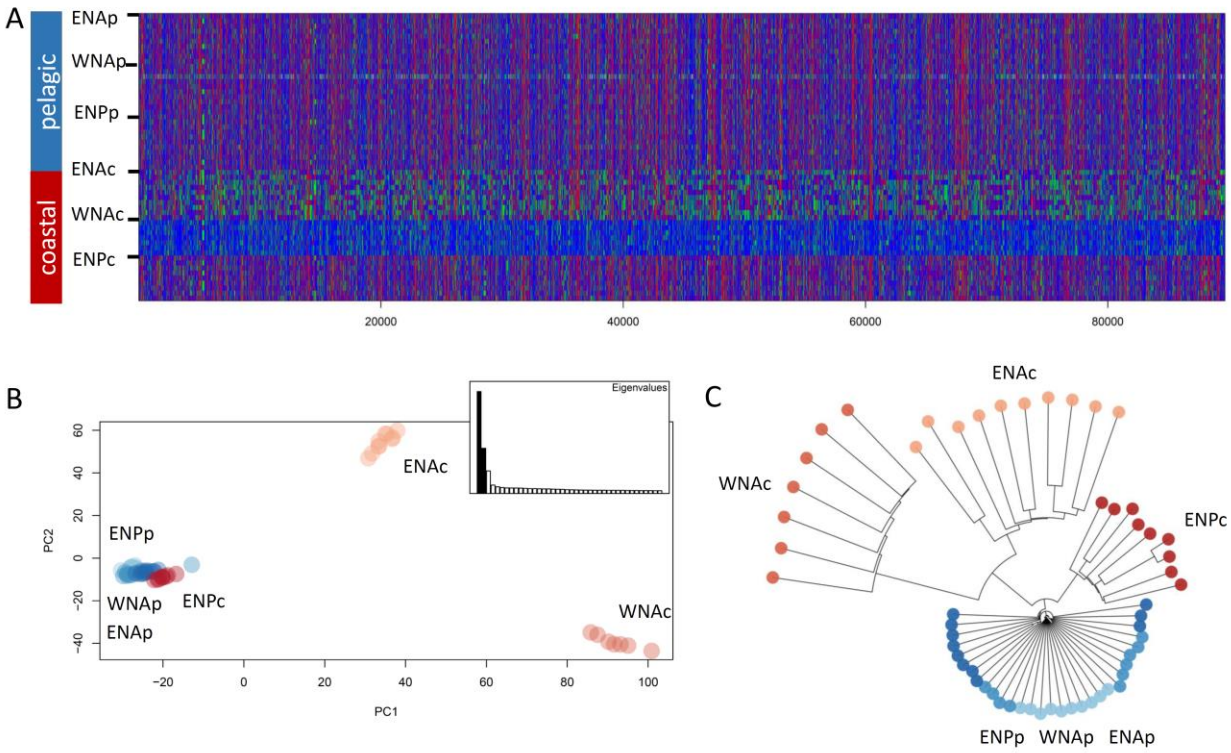

Fig. S17. Patterns of genetic variation of the 89,663 SNPs, scattered across the genome, under divergent selection between ecotypes, including closely linked SNPs scattered across the genome in 3,572 regions separated by at least 100 kb. (A) Plot of the genotypes, with blue: homozygote reference, green: heterozygote, and red: homozygote derived, (B) Principal component analysis and (C) Neighbor-joining distance tree of the common bottlenose dolphin samples.

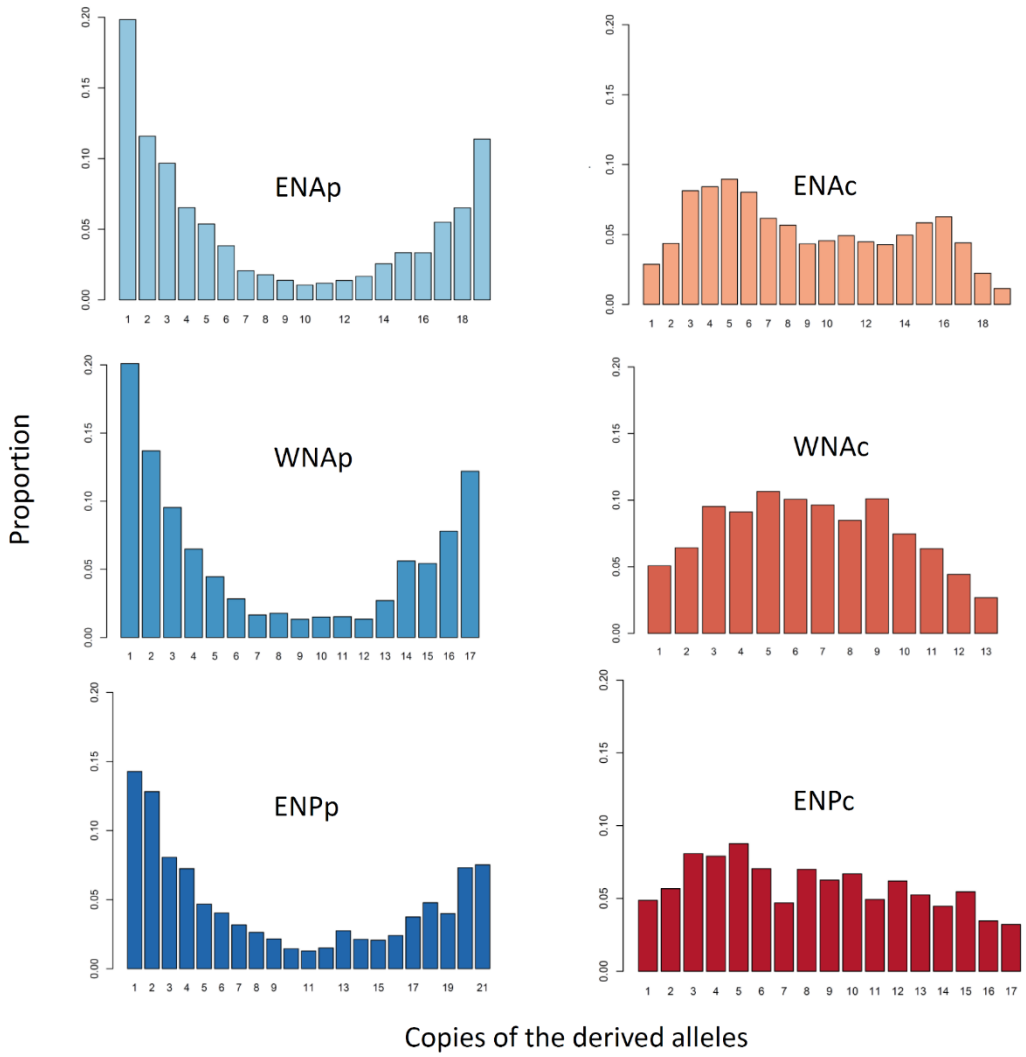

Fig. S18. Site frequency spectrum for the 7,165 SNPs under parallel linked selection for each of the bottlenose dolphin population, as the proportion of each number of copies of the derived alleles.

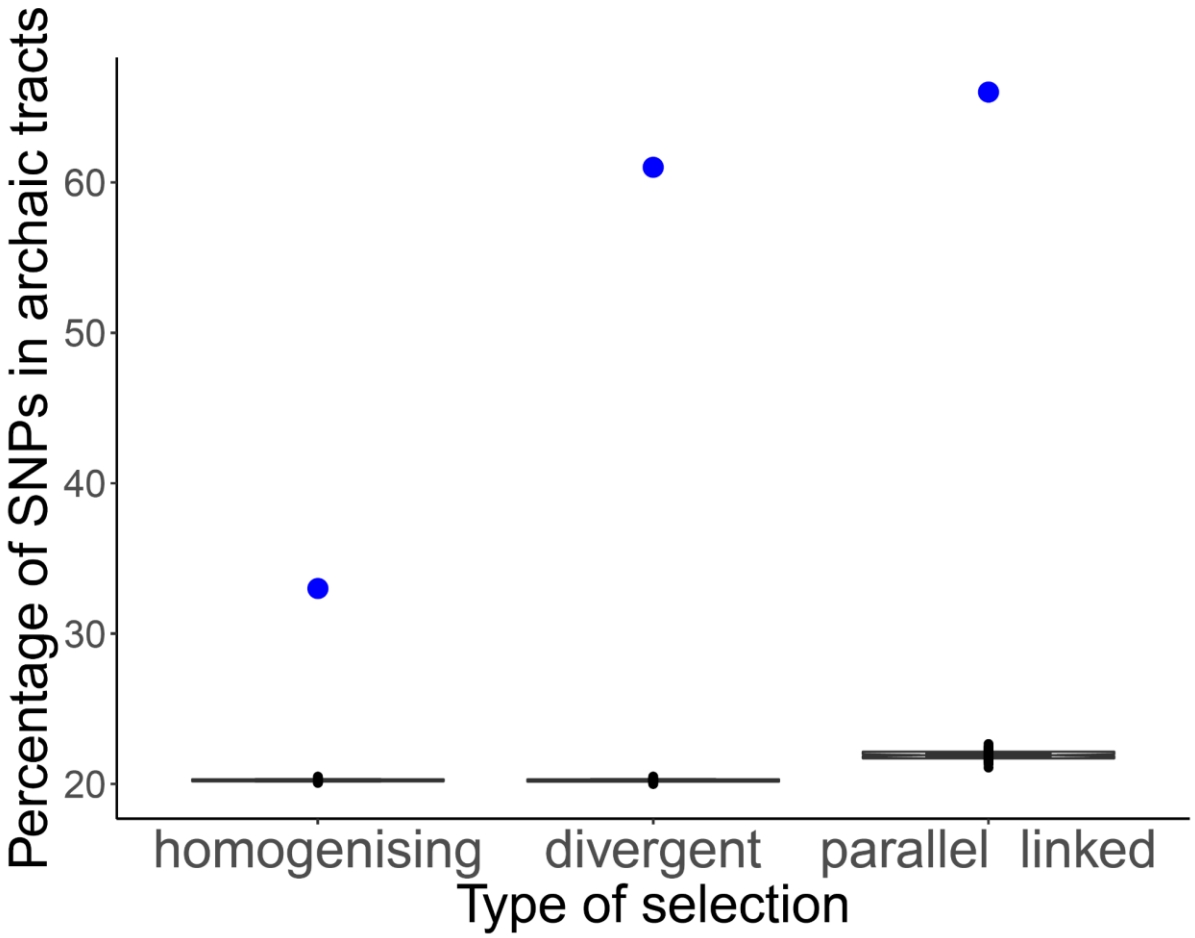

Fig. S19. Percentage of SNPs in ancient tracts under the different types of selection: i) homogenising selection among coastal populations "homogenising", ii) divergent selection between ecotypes "divergent" and iii) in "parallel linked" selection, that is, both under homogenizing selection among coastal populations and divergent selection between ecotypes, with the blue dots representing the observed percentage in our data and the boxplots the percentage in 100 random sample sets of the same number of putatively neutral SNPs.

Supplementary tables

Table S1. Sample information indicating the sample ID used in the laboratory (Lab_ID), sample ID used at the institution where the sample was taken (Institute_ID), population, sex, sampling date, number of raw reads after sequencing, mean coverage after filtering that is mean coverage of the bam files estimated using ANGSD, that is after the repeats, excessive coverage, mapping quality and phred score filters, mean coverage vcf, that is mean coverage of the filtered variants in the vcf file estimated with vcftools, Biosample and SRA Accession numbers.

Lab_ID	Institute_ID	Population	Sex	Sampling date	Number of raw reads	mean coverage after filtering	mean coverage vcf	Biosample number	SRA Accession number
S27	SW1992/201c	ENAc	M	10/11/1992	272014714	8.73	12.26	SAMN18839466	SRR14668022
S31	SW2011/173	ENAc	M	4/23/2011	254077360	8.75	12.47	SAMN18839467	SRR14668021
S37	SW1995/145a	ENAc	M	12/31/1995	237369566	8.44	11.94	SAMN18839468	SRR14668010
S38	SW1995/99a	ENAc	F	8/6/1995	260063378	8.52	12.01	SAMN18839469	SRR14667999
S39	SW1993/11c	ENAc	M	1/25/1993	231416096	7.78	11.11	SAMN18839470	SRR14667988
S41	SW1996/103c	ENAc	M	6/17/1996	250014606	8.66	12.11	SAMN18839471	SRR14667977
S42	SW2004/257a	ENAc	M	8/20/2004	232352130	7.53	11.0	SAMN18839472	SRR14667969
S53	SW1999/66b	ENAc	F	3/30/1999	256252676	8.78	12.32	SAMN18839473	SRR14667968
S54	SW1999/136A	ENAc	F	7/12/1999	265370600	8.80	12.47	SAMN18839474	SRR14667967
S56	SW2001/111a	ENAc	F	5/23/2001	244673284	8.23	11.08	SAMN18839475	SRR14667966
IR23	2007.1.179	ENAp	M	12/07/2009	282095448	9.13	12.93	SAMN18839476	SRR14668020
IR24	2007.1.180	ENAp	M	21/07/2009	278318284	9.34	13.22	SAMN18839477	SRR14668019
IR30	2007.1.276	ENAp	M	10/04/2012	236082662	8.63	11.84	SAMN18839478	SRR14668018

IR33	2007.1.273	ENAp	F	17/05/2011	309762602	10.46	14.55	SAMN18839479	SRR14668017
IR34	2007.1.272	ENAp	F	27/01/2012	231376318	8.07	10.79	SAMN18839480	SRR14668016
IR36	2007.1.270	ENAp	M	03/06/2011	260728180	8.80	12.34	SAMN18839481	SRR14668015
S10	SW2001/75a	ENAp	M	4/2/2001	337738358	11.32	16.57	SAMN18839482	SRR14668014
S12	SW1998/18a	ENAp	F	1/25/1998	235381406	7.75	10.85	SAMN18839483	SRR14668013
S40	SW2007/4c	ENAp	M	1/6/2007	298305240	9.95	14.51	SAMN18839484	SRR14668012
S43	SW2011/188	ENAp	M	5/1/2011	227419788	7.94	11.26	SAMN18839485	SRR14668011
115669	Tt1	ENPc	M	26/6/2013	270809186	6.53	8.36	SAMN18839486	SRR14668009
117692	CTTSD131023.01	ENPc	F	23/10/2013	267379282	6.17	8.00	SAMN18839487	SRR14668008
125942	CTTSD100723.02	ENPc	M	23/7/2010	286417592	6.89	9.13	SAMN18839488	SRR14668007
146413	CTTSD131213.01	ENPc	M	13/12/2013	264961758	6.46	8.77	SAMN18839489	SRR14668006
146414	CTTSD131213.02	ENPc	M	13/12/2013	282157544	6.83	8.63	SAMN18839490	SRR14668005
146416	HYDE150113.01	ENPc	M	13/1/2015	272176732	6.53	8.39	SAMN18839491	SRR14668004
146419	HYDE150128.01	ENPc	F	28/1/2015	271565530	6.46	8.09	SAMN18839492	SRR14668003
160317	LSK151102.04	ENPc	M	2/11/2015	216981988	5.33	7.07	SAMN18839493	SRR14668002
92200	CTTSD091102.01	ENPc	M	2/11/2009	282378954	6.52	9.23	SAMN18839494	SRR14668001
113122	CTTSD121029.06	ENPp	M	29/10/2012	266083128	6.21	8.82	SAMN18839495	SRR14668000
117696	GCAMPBELL131102.02	ENPp	F	2/11/2013	245179228	6.01	7.61	SAMN18839496	SRR14667998
117698	GCAMPBELL131103.01	ENPp	M	3/11/2013	262410774	6.59	8.89	SAMN18839497	SRR14667997

125944	CTTSD100730.02	ENPp	M	30/7/2010	274668084	6.55	8.44	SAMN18839498	SRR14667996
145428	DSJ141023.15	ENPp	F	23/10/2014	289591618	6.59	9.49	SAMN18839499	SRR14667995
145429	DSJ141023.16	ENPp	M	23/10/2014	247400236	5.90	8.08	SAMN18839500	SRR14667994
145430	DSJ141023.17	ENPp	F?	23/10/2014	224105208	4.30	7.37	SAMN18839501	SRR14667993
160321	LSK151104.02	ENPp	F	4/11/2015	254454974	6.04	8.34	SAMN18839502	SRR14667992
160322	LSK151104.03	ENPp	M	4/11/2015	265724004	6.38	8.79	SAMN18839503	SRR14667991
160326	LSK151105.04	ENPp	F	5/11/2015	275175338	6.41	8.22	SAMN18839504	SRR14667990
160335	LSK151107.06	ENPp	F	7/11/2015	271266882	6.62	8.56	SAMN18839505	SRR14667989
14Tt001	173059	WNAc	M	14/3/2002	310465756	9.82	12.88	SAMN18839506	SRR14667987
14Tt004	173060	WNAc	F	14/3/2002	271870888	7.51	12.01	SAMN18839507	SRR14667986
14Tt006	173061	WNAc	M	14/3/2002	309447666	9.46	14.04	SAMN18839508	SRR14667985
14Tt021	173062	WNAc	M	16/3/2002	292191584	9.33	12.03	SAMN18839509	SRR14667984
14Tt022	173063	WNAc	M	16/3/2002	317306584	7.74	11.82	SAMN18839510	SRR14667983
14Tt023	173064	WNAc	M	16/3/2002	275133508	7.16	10.51	SAMN18839511	SRR14667982
14Tt025	173065	WNAc	F	16/3/2002	304170502	6.50	11.08	SAMN18839512	SRR14667981
7Tt156	173048	WNAp	F	10/8/1999	238633104	7.27	9.30	SAMN18839513	SRR14667980
7Tt161	173049	WNAp	M	11/8/1999	278592266	8.67	11.48	SAMN18839514	SRR14667979
7Tt182*	173051	WNAp	M	16/8/1999	222939912	2.62*	4.48	SAMN18839515	SRR14667978
7Tt193	173052	WNAp	M	14/8/1999	263546918	7.73	10.86	SAMN18839516	SRR14667976

7Tt270	173053	WNAp	M	17/3/2002	217855012	7.08	9.74	SAMN18839517	SRR14667975
7Tt278	173054	WNAp	M	1/4/2002	260697324	8.05	11.72	SAMN18839518	SRR14667974
7Tt282	173055	WNAp	M	2/4/2002	254492318	6.02	10.72	SAMN18839519	SRR14667973
7Tt284	173056	WNAp	F	2/4/2002	252067218	6.87	11.52	SAMN18839520	SRR14667972
7Tt287	173057	WNAp	F	3/4/2002	244155564	8.02	11.12	SAMN18839521	SRR14667971
7Tt350	173058	WNAp	M	13/7/2005	253335024	7.91	10.51	SAMN18839522	SRR14667970

* Due to its low coverage, this individual was excluded from analyses not based on allele frequencies

Table S2. Type of data used for each of the analyses.

Analyses	data type
NGSAdmix (*39*)	genotype likelihoods
PCAs (*38*) in PCAngsd	genotype likelihoods
TreeMix v. 1.13 (*40*)	called genotypes - alleles frequencies
F_4-statistics (*41*, *42*) in TreeMix	called genotypes - alleles frequencies
SMC++ v. 1 (*46*)	called genotypes - genotypes and alleles frequencies
SFS and 2D-SFS in ANGSD v. 0.921 (*73*)	genotype likelihoods
diversity statistics in ANGSD v. 0.921	genotype likelihoods
F_{ST} estimates in vcftools v. 0.1.16 (*78*)	called genotypes - alleles frequencies
Ancient tracts (*53*)	pseudo-haploid random call
localPCA (*50*)	called genotypes
Flink (*47*)	called genotypes - alleles frequencies

Table S3. Mean pairwise F_{ST} as estimated in vcftools across all SNPs and standard deviation (SD).

F_{ST}	ENPc	ENPp	WNAc	WNAp	ENAc	ENAp
ENPc		0.10 (0.15)	0.32 (0.29)	0.13 (0.17)	0.24 (0.24)	0.13 (0.17)
ENPp			0.17 (0.24)	0.03 (0.10)	0.14 (0.18)	0.03 (0.09)
WNAc				0.18 (0.23)	0.26 (0.27)	0.16 (0.23)
WNAp					0.14 (0.18)	0.02 (0.08)
ENAc						0.12 (0.16)
ENAp						

Table S4. Number of regions identified as ancient tracts with posterior probabilities of >0.8 and their total length as estimated following Skov et al. 2018 (*53*) (see fig. S11).

Individual	number of regions with P>0.8	length in bp
ENAc1	1,464	15,487,000
ENAc2	1,172	14,910,000
ENAc3	1,136	14,380,000
ENAc4	1,174	15,209,000
ENAc5	1,243	13,937,000
ENAc6	1,210	12,362,000
ENAc7	1,207	15,273,000
ENAc8	1,339	14,320,000
ENAc9	1,241	14,778,000
ENAc10	1,141	14,725,000
WNAc1	2,067	24,861,000
WNAc2	1,755	22,939,000
WNAc3	1,982	20,528,000
WNAc4	2,097	23,085,000
WNAc5	2,037	23,645,000
WNAc6	1,853	21,869,000
WNAc7	2,108	15,377,000
ENPc1	1,370	11,608,000
ENPc2	1,373	11,867,000
ENPc3	1,537	12,400,000
ENPc4	1,500	10,666,000
ENPc5	1,496	11,161,000
ENPc6	1,330	11,429,000
ENPc7	1,372	11,858,000
ENPc8	1,678	10,185,000
ENPc9	1,496	12,714,000

Table S5. TMRCA between the coastal individual and the allopatric pelagic individual ($T_{Ingroup}$) and between the introgressed tracts within coastal dolphins and the corresponding genomic regions in the outgroup ($T_{Ancient}$, fig. S11) using mutation rates 1: 1.92e-8 substitution per nucleotide per generation (*82*), and 2: 2.56e-8 substitution per nucleotide per generation (*83*) in generations and in years using a generation time of 21.1 years (*81*).

	GENERATIONS		YEARS	
	Tingroup	***T***ancient	***T***ingroup	***T***ancient
mutation 1				
ENAc1	9,688	68,521	204,425	1,445,793
ENAc2	9,384	63,602	198,007	1,342,000
ENAc3	9,559	61,874	201,687	1,305,552
ENAc4	9,032	62,931	190,578	1,327,853
ENAc5	9,447	70,583	199,336	1,489,297
ENAc6	9,296	71,259	196,155	1,503,556
ENAc7	9,066	62,087	191,284	1,310,043
ENAc8	9,373	68,912	197,760	1,454,047
ENAc9	9,123	64,318	192,505	1,357,103
ENAc10	11,302	67,735	238,465	1,429,215
WNAc1	15,564	104,506	328,405	2,205,080
WNAc2	18,583	107,003	392,108	2,257,754
WNAc3	16,676	102,630	351,859	2,165,498
WNAc4	15,004	106,478	316,580	2,246,687
WNAc5	15,364	100,307	324,190	2,116,469
WNAc6	17,498	105,878	369,203	2,234,025
WNAc7	20,115	107,629	424,417	2,270,966
ENPc1	10,311	76,607	217,570	1,616,405
ENPc2	10,366	76,303	218,723	1,610,002
ENPc3	10,610	80,668	223,869	1,702,089
ENPc4	10,532	81,946	222,222	1,729,066
ENPc5	10,352	79,208	218,437	1,671,298
ENPc6	10,312	75,117	217,582	1,584,968
ENPc7	10,454	78,553	220,582	1,657,478
ENPc8	10,677	80,468	225,279	1,697,869
ENPc9	10,380	76,218	219,028	1,608,199
mutation 2				
ENAc1	4,592	32,474	96,884	685,210
ENAc2	4,448	30,143	93,842	636,019
ENAc3	4,530	29,324	95,586	618,745
ENAc4	4,281	29,825	90,322	629,314
ENAc5	4,477	33,452	94,472	705,828
ENAc6	4,406	33,772	92,964	712,586
ENAc7	4,296	29,425	90,656	620,874
ENAc8	4,442	32,660	93,725	689,122

ENAc9	4,324	30,482	91,234	643,177
ENAc10	5,356	32,102	113,017	677,353
WNAc1	7,376	49,529	155,642	1,045,062
WNAc2	8,807	50,712	185,833	1,070,026
WNAc3	7,903	48,640	166,758	1,026,302
WNAc4	7,111	50,464	150,038	1,064,781
WNAc5	7,282	47,539	153,645	1,003,066
WNAc6	8,293	50,179	174,978	1,058,780
WNAc7	9,533	51,009	201,146	1,076,287
ENPc1	4,887	36,307	103,114	766,069
ENPc2	4,913	36,163	103,660	763,034
ENPc3	5,028	38,231	106,099	806,677
ENPc4	4,991	38,837	105,319	819,463
ENPc5	4,906	37,540	103,524	792,084
ENPc6	4,887	35,600	103,120	751,170
ENPc7	4,955	37,229	104,541	785,535
ENPc8	5,060	38,136	106,767	804,677

Table S6. Percentage of shared ancient tracts between pairs of individuals, as the percentage of ancient tracts in individual 1 (i1) that are ancient as well in individual 2 (i2, above diagonal), and the percentage of ancient tracts in individual 2 that are also ancient in individual 1 (below diagonal). Due to the action of recombination, we do not expect individuals within a population to share 100% of ancient tracts, as windows are small (1 kb) and we only kept those with a certain posterior probability. In addition, it is not expected, for instance there are few examples of fixed or high frequency ancient hominin tracts within any given modern human population.

i1\ i2	ENAc1	ENAc2	ENAc3	ENAc4	ENAc5	ENAc6	ENAc7	ENAc8	ENAc9	ENAc1	WNAc	WNAc	WNAc	WNAc	WNAc	WNAc	ENPc1	ENPc2	ENPc3	ENPc4	ENPc5	ENPc6	ENPc7	ENPc8	ENPc9
ENAc1		32.8	30.4	29.1	26.1	27	32.7	32.2	34.3	37.6	9.5	9.2	8.9	8.7	9.6	9.3	8.1	2.3	2.2	2.3	2.2	2.3	2.1	2.3	2.2	2.6
ENAc2	34.4		35.9	32.7	31.5	28.2	34.1	31.4	32.7	30.7	10.1	9.5	8.9	9.4	9.9	9.2	8.2	1.9	2.2	2.1	2.2	2.1	2	2	1.9	2.6
ENAc3	30.1	33.8		31.6	29.9	29.9	31.2	29.7	33.8	27.8	10.1	9.9	9.4	9.4	10	9.6	8.7	2.1	2.4	2.4	2.4	2.2	2.4	2.2	2.1	2.6
ENAc4	29.6	31.7	32.5		31.1	29.2	34.4	29.5	33.7	30.4	10.8	11.3	10.3	9.9	11.2	10.5	9	2.1	2.3	2.4	2.2	2.2	2.3	2.1	2.1	2.8
ENAc5	30.5	35.1	35.3	35.8		32.7	39.8	35.4	35.1	27.6	12	11.5	10.7	10.7	12	11.3	9.6	2.2	2.3	2.4	2.6	2.3	2.4	2.3	2.1	2.6
ENAc6	34	35.1	39.4	37.4	36.4		36.7	33.7	35.4	31.5	12.1	11.7	11.3	11.4	12.4	11.5	10	1.9	3	2.5	2.2	2.3	2.4	2.7	2.2	2.7
ENAc7	32.4	32.2	31.3	33.5	33.7	27.8		32.6	32.7	27.5	9.7	9.9	9.1	9.7	9.9	9.5	8.5	2.1	2.1	2.3	2.1	2.3	2	2.1	2.2	2.6
ENAc8	36.7	34.1	34.3	33.1	34.5	29.5	37.5		46.3	31.7	10.7	10.6	10.1	10.5	10.8	10.6	9.2	2.1	2.2	2.4	2.1	2.4	2.4	2.4	2.1	2.8
ENAc9	35.6	32.4	33.9	34.5	31.1	28.2	34.3	42.2		31.4	10.4	10.1	9.6	9.6	10.3	9.9	8.8	2	2.5	2.3	2.3	2.4	2	2.4	2.3	2.7
ENAc10	37.7	29.4	28.2	30	23.6	24.3	27.9	27.9	30.3		9.6	9.4	9.5	9.3	9.6	9.6	8.5	2.3	2.5	2.3	2.2	2.1	2.3	2.4	2.3	2.5
WNAc1	7.6	7.7	8.2	8.5	8.2	7.4	7.9	7.6	8	7.7		51.8	49.1	52	53.7	49.9	40.8	2	2.1	2.1	2	2	2.1	2.2	2.1	2.3
WNAc2	7.5	7.4	8.1	9.1	8	7.4	8.1	7.6	7.9	7.7	52.6		48.3	50.7	52.4	48.6	39.8	2	2.2	2.2	2	2.1	2.1	2.3	2.3	2.5
WNAc3	7.2	6.9	7.8	8.3	7.5	7.1	7.5	7.2	7.5	7.7	50	48.3		49	49.8	47.1	40.4	2.1	2.1	2.2	2.2	2.1	2.1	2.4	2.2	2.4
WNAc4	7.6	7.8	8.3	8.4	8	7.6	8.5	8.1	8	8.1	56.3	54	52.2		56.3	51.8	42.9	2	1.9	2	1.9	1.8	1.9	2	2.2	2.3
WNAc5	7.5	7.3	7.9	8.5	7.9	7.4	7.7	7.4	7.7	7.4	51.7	49.7	47.2	50.1		47.7	40	1.9	2.1	2.1	1.9	2	2	2.2	2.1	2.3
WNAc6	8.2	7.7	8.5	9.1	8.4	7.7	8.4	8.2	8.4	8.4	54.4	52	50.4	52.1	53.9		42.1	2	1.9	2.2	2	2	2.1	2.3	2	2.4
WNAc7	7	6.8	7.5	7.6	7.1	6.6	7.4	6.9	7.3	7.3	43.7	42	42.6	42.5	44.5	41.4		2	2.1	2.3	2.2	2	2.1	2.3	2.1	2.5
ENPc1	3	2.5	2.9	2.8	2.5	2	2.8	2.5	2.6	3.1	3.3	3.3	3.5	3.1	3.4	3	3.1		36.1	36.4	35.5	35.7	36.4	37.9	40.6	37.9
ENPc2	2.9	2.7	3.1	2.9	2.5	2.9	2.7	2.5	3.1	3.2	3.4	3.4	3.4	2.8	3.4	2.8	3.2	34.5		37.3	33.5	35.2	34.8	35.6	33.1	37.9
ENPc3	2.9	2.5	3	3	2.6	2.4	2.9	2.6	2.8	2.9	3.4	3.4	3.4	2.9	3.5	3.2	3.4	34.5	37		34.1	34.4	40.6	37.5	32.3	58.3
ENPc4	3.1	3	3.4	3	3.1	2.3	3	2.7	3.2	3.1	3.5	3.5	3.8	3.1	3.5	3.2	3.6	37.5	37.1	38		35.9	36.5	37.9	35.3	38.9
ENPc5	3.2	2.9	3.1	3.1	2.8	2.5	3.3	3	3.2	2.9	3.5	3.6	3.7	3	3.6	3.2	3.2	37.8	38.9	38.3	35.9		36.6	37.8	33.8	39.3
ENPc6	2.7	2.4	3.1	2.9	2.6	2.3	2.6	2.7	2.5	2.9	3.3	3.3	3.3	2.8	3.2	3	3.1	34.6	34.6	40.6	32.8	32.9		35.7	31.9	42.2
ENPc7	2.9	2.4	2.9	2.6	2.5	2.6	2.7	2.7	3	3.1	3.4	3.5	3.7	2.9	3.6	3.4	3.4	36	35.4	37.6	34.1	33.9	35.7		32.4	38.8
ENPc8	3	2.5	3	2.8	2.5	2.3	3	2.5	3	3.2	3.5	3.8	3.6	3.5	3.7	3.2	3.3	41.3	35.2	34.7	34	32.5	34.2	34.7		35.3
ENPc9	3.1	3	3.2	3.3	2.6	2.4	3.1	2.9	3.1	3	3.4	3.6	3.5	3.2	3.6	3.2	3.4	33.7	35.2	54.5	32.7	33	39.5	36.3	30.8	

Table S7. List of the 45 genes associated with the SNPs under parallel linked selection (i.e. under both homogenising selection in the coastal populations and divergent between ecotypes showing both 1% FDR threshold, above which we consider the locus under selection), and putative functions

Genes	Other gene names	Putative functions
RPL10A		Encodes a ribosomal protein that is a component of the 60S subunit of the ribosomes
TEAD3		Encodes the transcription factor TEF, which plays a role in the Hippo signaling pathway, a pathway involved in organ size control and tumor suppression
GCM1		Encodes a transcription factor involved in the control of expression of placental growth factor (PGF) and other placenta-specific genes
LOC101328860	HLA-DQA2	Encodes the HLA class II histocompatibility antigen, DQ alpha 2 chain, of immunologic importance
LOC109552651	ST3GAL1	Encodes CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 2-like. Cell type-specific expression of unique carbohydrate structures on cell surface glycoproteins and glycolipids provides information relevant to cell-cell interactions in developing and adult organisms.
LPIN2		Encodes phosphatidate phosphatase LPIN2. Has a role in controlling the metabolism of fatty acids, and body composition, such as fat-mass ratio (*65*).
LIAS		Encodes lipoyl synthase, involved in lipoic acid synthase, plays a central role in the antioxidant network
CAMK2D		Encodes a calcium/calmodulin-dependent protein kinase involved in the regulation of Ca(2+) homeostatis and excitation-contraction coupling (ECC). Targets also transcription factors and signaling molecules to regulate heart function
LOC101323008	MYH13-like	Encodes myosin-13 which is involved in muscle contraction
LOC101324002	MH3	Encodes myosin-3, involved in heart contraction
ADARB2	ADAR3	Encodes double-stranded RNA-specific editase B2. Has a role in RNA editing, and involve in cognitive, learning and memory abilities (*63*).
MMP20		Encodes enamelysin, which has a role in dental enamel formation
LOC109552895	putative dimethylaniline monooxygenase [N-oxide-forming] 6	COQ6 is a flavin-dependent monooxygenase needed for biosynthesis of coenzyme Q10. Coenzyme Q10 has a role as redox carrier in the mitochondrial respiratory chain and as a lipid-soluble antioxidant implicated in protection from cell damage by reactive oxygen species.
LOC101318787	G protein-coupled receptor 89A	Voltage dependent anion channel needed for acidification and functions of the Golgi apparatus
ZNF697		Encodes the Zinc finger protein 697 and may be involved in transcriptional regulation

DPYD		Encodes dihydropyrimidine dehydrogenase, involved in pyrimidine base degradation. Catalyzes the reduction of uracil and thymine. Also involved the degradation of the chemotherapeutic drug 5-fluorouracil. Involved in β-Alanine production, a putative neurotransmitter and a component of a number of coenzyme A and endogenous antioxidants found in the brain.
STAR		Encode the steroidogenic Acute Regulatory Protein, involved in acute regulation of steroid hormone synthesis by enhancing the conversion of cholesterol into pregnenolone
PALB2		Encodes Partner And Localizer Of BRCA2, involved in maintenance of genome stability, specifically the homologous recombination pathway for double-strand DNA repair
MCM3AP		The protein encoded by this gene is a MCM3 binding protein, which are essential for the initiation of DNA replication
RIOX2		Encodes the Ribosomal Oxygenase 2, which regulates immune responses and may play an important role in cell growth and survival
GABRR3		The neurotransmitter gamma-aminobutyric acid (GABA) regulates synaptic transmission of neurons in the central nervous system. GABRR3 encodes one of three related subunits of the gene. Also has a potential role in aging and longevity.
CEP295		Encodes a centriole-enriched microtubule-binding protein with a role in elongating procentrioles after formation of the initiating cartwheel hub andposttranslational modification of centriolar microtubules
TAF1D		TAF1D is part of the SL1 complex which has a role in RNA polymerase I transcription
CUNH11orf54		Encodes ester hydrolase C11orf54 homolog which shows ester hydrolase activity on the substrate p-nitrophenyl acetate
LOC101334760	AGK – acylglycerol kinase	Encodes a lipid kinase involved in lipid and glycerolipid metabolism, catalyzes the formation of phosphatidic and lysophosphatidic acids
CEP152		Encodes a centrosomal protein involved with centrosome function, has a role in cell shape, polarity, motility, and division
LOC101331561	GPR133, ADGRD1 adhesion G protein-coupled receptor D1	Encodes a membrane-bound protein with long N termini containing multiple domains, possibly associated with adult height
LOC101337631	CYB5R4	Encodes the cytochrome b5 reductase 4 involved in endoplasmic reticulum stress response pathway, protection of pancreatic beta-cells against oxidant stress and in in vitro reduction of cytochrome c, feericyanide and methemoglobin

LOC101316550	CCDC162P (coiled-coil domain-containing protein 162)	Transcripts from this locus encode truncated proteins, and may be involved in nonsense-mediated decay
SERINC5		Encodes serine incorporator, which enhances the incorporation of serine into phosphatidylserine and sphingolipids, restrict infectivity of lentiviruses
CCDC93		Protein coding gene, component of the CCC complex, which has a role in the regulation of endosomal recycling of surface proteins, including integrins, signaling receptor and channels
ACER2		The ceramidase ACER2 hydrolyzes very long chain ceramides to generate sphingosine. Ceramides and sphingosine are bioactive lipids mediating cellular signaling pathways
AAMP		Encodes a protein associated with angiogenesis and cell migration
GPBAR1	TGR5	Encodes a member of the G protein-coupled receptor (GPCR) superfamily, which functions as a cell surface receptor for bile acids
FAM196A	INSYN2A	Component of the protein machinery at the inhibitory synapses. This synaptic inhibition is central to the functioning of the central nervous system, it shapes and orchestrates the flow of information through neuronal networks to produce a precise neural code.
PAOX		Encodes a Flavoenzyme which catalyzes the oxidation of N(1)-acetylspermine to spermidine and is involved in the polyamine back-conversion
MTG1		Has a role in the regulation of the mitochondrial ribosome assembly and of translational activity, and shows mitochondrial GTPase activity
RELN		Encodes the reelin protein, which has a role in the modulation of synaptic transmission in response to experience, learning and memory (*61*, *62*).
RYR1		Encodes a ryanodine receptor found in skeletal muscle. The encoded protein functions as a calcium release channel in the sarcoplasmic reticulum, thereby have a key role in triggering muscle contraction and body movement. Can also mediate the release of calcium from intracellular stores in neurons.
PATE2		PATE2 (Prostate And Testis Expressed 2) is a protein coding gene
FEZ1		Encodes the Fasciculation and elongation protein zeta-1, which may be involved in axonal outgrowth
FEZ2		Encodes the Fasciculation and elongation protein zeta-1, which is involved in axonal outgrowth and fasciculation
CCDC57		Coiled-Coil Domain Containing 57 is a protein coding gene. Centrosomes function in key cellular processes ranging from cell division to cellular signaling

SLC17A3		The protein encoded by this gene is a voltage-driven transporter that excretes intracellular urate and organic anions from the blood into renal tubule cells
LOC101332501	ZNF501	Encodes the Zinc Finger Protein 501, which may be involved in transcriptional regulation

REFERENCES AND NOTES

1. T. J. Kawecki, D. Ebert, Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).
2. J. Li, H. Li, M. Jakobsson, S. Li, P. Sjödin, M. Lascoux, Joint analysis of demography and selection in population genetics: Where do we stand and where could we go? *Mol. Ecol.* **21**, 28–44 (2012).
3. B. Charlesworth, M. Nordborg, D. Charlesworth, The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70**, 155–174 (1997).
4. G. L. Conte, M. E. Arnegard, C. L. Peichel, D. Schluter, The probability of genetic parallelism and convergence in natural populations. *Proc. Biol. Sci.* **279**, 5039–5047 (2012).
5. F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, D. M. Kingsley, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
6. V. Soria-Carrasco, Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, P. Nosil, Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**, 738–742 (2014).
7. M. Foll, O. E. Gaggiotti, J. T. Daub, A. Vatsiou, L. Excoffier, Widespread signals of convergent adaptation to high altitude in Asia and America. *Am. J. Hum. Genet.* **95**, 394–407 (2014).
8. O. Seehausen, Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. J. Mrosso, R. Miyagi, I. van der Sluijs, M. V. Schneider, M. E. Maan, H. Tachida, H. Imai, N. Okada, Speciation through sensory drive in cichlid fish. *Nature* **455**, 620–626 (2008).
9. J. M. Alves, M. Carneiro, J. Y. Cheng, A. Lemos de Matos, M. M. Rahman, L. Loog, P. F. Campos, N. Wales, A. Eriksson, A. Manica, T. Strive, S. C. Graham, S. Afonso, D. J. Bell, L. Belmont, J. P. Day, S. J. Fuller, S. Marchandeau, W. J. Palmer, G. Queney, A. K. Surridge, F. G. Vieira, G. McFadden, R. Nielsen, M. T. P. Gilbert, P. J. Esteves, N. Ferrand, F. M. Jiggins, Parallel adaptation of rabbit populations to myxoma virus. *Science* **363**, 1319–1326 (2019).
10. K. A. Thompson, M. M. Osmond, D. Schluter, Parallel genetic evolution and speciation from standing variation. *Evol. Lett.* **3**, 129–141 (2019).

11. D. I. Bolnick, R. D. H. Barrett, K. B. Oke, D. J. Rennison, Y. E. Stuart, (Non)parallel evolution. *Annu. Rev. Ecol. Evol. Syst.* **49**, 303–330 (2018).
12. Y. E. Stuart, T. Veen, J. N. Weber, D. Hanson, M. Ravinet, B. K. Lohman, C. J. Thompson, T. Tasneem, A. Doggett, R. Izen, N. Ahmed, R. D. H. Barrett, A. P. Hendry, C. L. Peichel, D. I. Bolnick, Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat. Ecol. Evol.* **1**, 0158 (2017).
13. B. Fang, P. Kemppainen, P. Momigliano, X. Feng, J. Merilä, On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat. Ecol. Evol.* **4**, 1105–1115 (2020).
14. S. P. De Lisle, D. I. Bolnick, A multivariate view of parallel evolution. *Evolution* **74**, 1466–1481 (2020).
15. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
16. S. M. Van Belleghem, C. Vangestel, K. De Wolf, Z. De Corte, M. Möst, P. Rastas, L. De Meester, F. Hendrickx, Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLOS Genet.* **14**, e1007796 (2018).
17. R. F. Guerrero, M. W. Hahn, Speciation as a sieve for ancestral polymorphism. *Mol. Ecol.* **26**, 5362–5368 (2017).
18. Q. Rougemont, P.-A. Gagnaire, C. Perrier, C. Genthon, A.-L. Besnard, S. Launey, G. Evanno, Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol. Ecol.* **26**, 142–162 (2017).
19. D. A. Marques, J. I. Meier, O. Seehausen, A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019).
20. T. C. Nelson, W. A. Cresko, Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.* **2**, 9–21 (2018).
21. S. A. Tishkoff, F. A. Reed, A. Ranciaro, B. F. Voight, C. C. Babbitt, J. S. Silverman, K. Powell, H. M. Mortensen, J. B. Hirbo, M. Osman, M. Ibrahim, S. A. Omar, G. Lema, T. B. Nyambo, J. Ghori, S. Bumpstead, J. K. Pritchard, G. A. Wray, P. Deloukas, Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).
22. B. Charlesworth, Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).

23. A. D. Foote, N. Vijay, M. C. Ávila-Arcos, R. W. Baird, J. W. Durban, M. Fumagalli, R. A. Gibbs, M. B. Hanson, T. S. Korneliussen, M. D. Martin, K. M. Robertson, V. C. Sousa, F. G. Vieira, T. Vinař, P. Wade, K. C. Worley, L. Excoffier, P. A. Morin, M. T. P. Gilbert, J. B. W. Wolf, Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat. Commun.* **7**, 11693 (2016).
24. A. R. Hoelzel, C. W. Potter, P. B. Best, Genetic differentiation between parapatric "nearshore" and "offshore" populations of the bottlenose dolphin. *Proc. Biol. Sci.* **265**, 1177–1183 (1998).
25. A. Natoli, V. M. Peddemors, A. R. Hoelzel, Population structure and speciation in the genus *Tursiops* based on microsatellite and mitochondrial DNA analyses. *J. Evol. Biol.* **17**, 363–375 (2004).
26. M. Louis, A. Viricel, T. Lucas, H. Peltier, E. Alfonsi, S. Berrow, A. Brownlow, P. Covelo, W. Dabin, R. Deaville, R. de Stephanis, F. Gally, P. Gauffier, R. Penrose, M. A. Silva, C. Guinet, B. Simon-Bouhet, Habitat-driven population structure of bottlenose dolphins, *Tursiops truncatus*, in the North-East Atlantic. *Mol. Ecol.* **23**, 857–874 (2014).
27. A. P. B. Costa, P. F. Fruet, E. R. Secchi, F. G. Daura-Jorge, P. C. Simões-Lopes, J. C. Di Tullio, P. E. Rosel, Ecological divergence and speciation in common bottlenose dolphins in the western South Atlantic. *J. Evol. Biol.* **34**, 16–32 (2019).
28. J. L. Lowther-Thieleking, F. I. Archer, A. R. Lang, D. W. Weller, Genetic differentiation among coastal and offshore common bottlenose dolphins, *Tursiops truncatus*, in the eastern North Pacific Ocean. *Mar. Mamm. Sci.* **31**, 1–20 (2015).
29. M. Nykänen, K. Kaschner, W. Dabin, A. Brownlow, N. J. Davison, R. Deaville, C. Garilao, K. Kesner-Reyes, M. T. P. Gilbert, R. Penrose, V. Islas-Villanueva, N. Wales, S. N. Ingram, E. Rogan, M. Louis, A. D. Foote, Postglacial colonization of northern coastal habitat by bottlenose dolphins: A marine leading-edge expansion? *J. Hered.* **110**, 662–674 (2019).
30. M. Louis, M. C. Fontaine, J. Spitz, E. Schlund, W. Dabin, R. Deaville, F. Caurant, Y. Cherel, C. Guinet, B. Simon-Bouhet, Ecological opportunities and specializations shaped genetic divergence in a highly mobile marine top predator. *Proc. Biol. Sci.* **281**, 28120141558 (2014).
31. J. G. Mead, C. W. Potter, Recognizing two populations of the bottlenose dolphin (*Tursiops truncatus*) of the Atlantic coast of North America-morphologic and ecologic considerations (1995); https://repository.si.edu/bitstream/handle/10088/2652/mead_and_potter1995.pdf?sequence=1.

32. A. P. B. Costa, P. E. Rosel, F. G. Daura-Jorge, P. C. Simões-Lopes, Offshore and coastal common bottlenose dolphins of the western South Atlantic face-to-face: What the skull and the spine can tell us. *Mar. Mamm. Sci.* **32**, 1433–1457 (2016).
33. W. F. Perrin, J. L. Thieleking, W. A. Walker, F. I. Archer, K. M. Robertson, Common bottlenose dolphins (*Tursiops truncatus*) in California waters: Cranial differentiation of coastal and offshore ecotypes. *Mar. Mamm. Sci.* **27**, 769–792 (2011).
34. P. E. Rosel, L. Hansen, A. A. Hohn, Restricted dispersal in a continuously distributed marine species: Common bottlenose dolphins *Tursiops truncatus* in coastal waters of the western North Atlantic. *Mol. Ecol.* **18**, 5030–5045 (2009).
35. A. M. S. Machado, M. Cantor, A. P. B. Costa, B. P. H. Righetti, C. Bezamat, J. V. S. Valle-Pereira, P. C. Simões-Lopes, P. V. Castilho, F. G. Daura-Jorge, Homophily around specialized foraging underlies dolphin social preferences. *Biol. Lett.* **15**, 20180909 (2019).
36. A. M. Kopps, C. Y. Ackermann, W. B. Sherwin, S. J. Allen, L. Bejder, M. Krützen, Cultural transmission of tool use combined with habitat specializations leads to fine-scale genetic structure in bottlenose dolphins. *Proc. Biol. Sci.* **281**, 20133245 (2014).
37. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
38. J. Meisner, A. Albrechtsen, Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
39. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
40. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* **8**, e1002967 (2012).
41. D. Reich, K. Thangaraj, N. Patterson, A. L. Price, L. Singh, Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
42. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
43. M. Leitwein, M. Duranton, Q. Rougemont, P.-A. Gagnaire, L. Bernatchez, Using haplotype information for conservation genomics. *Trends Ecol. Evol.* **35**, 245–258 (2020).

44. O. Mazet, W. Rodríguez, S. Grusea, S. Boitard, L. Chikhi, On the importance of being structured: Instantaneous coalescence rates and human evolution—Lessons for ancestral population size inference? *Heredity* **116**, 362–371 (2016).
45. D. R. Schrider, A. G. Shanku, A. D. Kern, Effects of linked selective sweeps on demographic inference and model selection. *Genetics* **204**, 1207–1223 (2016).
46. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
47. M. Galimberti, C. Leuenberger, B. Wolf, S. M. Szilágyi, M. Foll, D. Wegmann, Detecting selection from linked sites using an *F*-model. *Genetics* **216**, 1205–1215 (2020).
48. R. Nielsen, Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**, 641–647 (2001).
49. H. Bailey, P. Thompson, Effect of oceanographic features on fine-scale foraging movements of bottlenose dolphins. *Mar. Ecol. Prog. Ser.* **418**, 223–233 (2010).
50. H. Li, P. Ralph, Local PCA shows how the effect of population structure differs along the genome. *Genetics* **211**, 289–304 (2019).
51. M. Foll, O. Gaggiotti, A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993 (2008).
52. W. Stephan, Selective sweeps. *Genetics* **211**, 5–13 (2019).
53. L. Skov, R. Hui, V. Shchur, A. Hobolth, A. Scally, M. H. Schierup, R. Durbin, Detecting archaic introgression using an unadmixed outgroup. *PLOS Genet.* **14**, e1007641 (2018).
54. A. E. Moura, K. Shreves, M. Pilot, K. R. Andrews, D. M. Moore, T. Kishida, L. Möller, A. Natoli, S. Gaspari, M. McGowen, I. Chen, H. Gray, M. Gore, R. M. Culloch, M. S. Kiani, M. S. Willson, A. Bulushi, T. Collins, R. Baldwin, A. Willson, G. Minton, L. Ponnampalam, A. R. Hoelzel, Phylogenomics of the genus *Tursiops* and closely related Delphininae reveals extensive reticulation among lineages and provides inference about eco-evolutionary drivers. *Mol. Phylogenet. Evol.* **146**, 106756 (2020).
55. M. R. McGowen, G. Tsagkogeorga, S. Álvarez-Carretero, M. Dos Reis, M. Struebig, R. Deaville, P. D. Jepson, S. Jarman, A. Polanowski, P. A. Morin, S. J. Rossiter, Phylogenomic resolution of the cetacean tree of life using target sequence capture. *Syst. Biol.* **69**, 479–501 (2020).

56. M. Duranton, F. Allal, S. Valière, O. Bouchez, F. Bonhomme, P.-A. Gagnaire, The contribution of ancient admixture to reproductive isolation between European sea bass lineages. *Evol. Lett.* **4**, 226–242 (2020).
57. A. D. Foote, M. D. Martin, M. Louis, G. Pacheco, K. M. Robertson, M. S. Sinding, A. R. Amaral, R. W. Baird, C. S. Baker, L. Ballance, J. Barlow, A. Brownlow, T. Collins, R. Constantine, W. Dabin, L. D. Rosa, N. J. Davison, J. W. Durban, R. Esteban, S. H. Ferguson, T. Gerrodette, C. Guinet, M. Bradley Hanson, W. Hoggard, C. J. D. Matthews, F. I. P. Samarra, R. de Stephanis, S. B. Tavares, P. Tixier, J. A. Totterdell, P. Wade, L. Excoffier, M. T. P. Gilbert, J. B. W. Wolf, P. A. Morin, Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Mol. Ecol.* **28**, 3427–3444 (2019).
58. P. F. Colosimo, K. E. Hosemann, S. Balabhadra, G. Villarreal Jr., M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, D. M. Kingsley, Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
59. D. Schluter, G. L. Conte, Genetics and ecological speciation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9955–9962 (2009).
60. J. K. Pritchard, J. K. Pickrell, G. Coop, The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010).
61. E. J. Weeber, U. Beffert, C. Jones, J. M. Christian, E. Forster, J. D. Sweatt, J. Herz, Reelin and ApoE receptors cooperate to enhance hippocampal synaptic plasticity and learning. *J. Biol. Chem.* **277**, 39944–39952 (2002).
62. J. Herz, Y. Chen, Reelin, lipoprotein receptors and synaptic plasticity. *Nat. Rev. Neurosci.* **7**, 850–859 (2006).
63. D. Mladenova, G. Barry, L. M. Konen, S. S. Pineda, B. Guennewig, L. Avesson, R. Zinn, N. Schonrock, M. Bitar, N. Jonkhout, L. Crumlish, D. C. Kaczorowski, A. Gong, M. Pinese, G. R. Franco, C. R. Walkley, B. Vissel, J. S. Mattick, Adar3 is involved in learning and memory in mice. *Front. Neurosci.* **12**, 243 (2018).
64. A. C. Beichman, K.-P. Koepfli, G. Li, W. Murphy, P. Dobrynin, S. Kliver, M. T. Tinker, M. J. Murray, J. Johnson, K. Lindblad-Toh, E. K. Karlsson, K. E. Lohmueller, R. K. Wayne, Aquatic adaptation and depleted diversity: A deep dive into the genomes of the sea otter and giant otter. *Mol. Biol. Evol.* **36**, 2631–2655 (2019).

65. Y. S. Aulchenko, J. Pullen, W. P. Kloosterman, M. Yazdanpanah, A. Hofman, N. Vaessen, P. J. L. M. Snijders, D. Zubakov, I. Mackay, M. Olavesen, B. Sidhu, V. E. Smith, A. Carey, E. Berezikov, A. G. Uitterlinden, R. H. A. Plasterk, B. A. Oostra, C. M. van Duijn, LPIN2 is associated with type 2 diabetes, glucose metabolism, and body composition. *Diabetes* **56**, 3020–3026 (2007).
66. R. A. Oomen, A. Kuparinen, J. A. Hutchings, Consequences of single-locus and tightly linked genomic architectures for evolutionary responses to environmental change. *J. Hered.* **111**, 319–332 (2020).
67. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
68. A. E. Moura, S. C. A. Nielsen, J. T. Vilstrup, J. V. Moreno-Mayar, M. T. P. Gilbert, H. W. I. Gray, A. Natoli, L. Möller, A. R. Hoelzel, Recent diversification of a marine genus (*Tursiops spp.*) tracks habitat preference and environmental change. *Syst. Biol.* **62**, 865–877 (2013).
69. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
70. Broad Institute, Picard tools (2016).
71. G. A. Van der Auwera, B. D. O'Connor, *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media Inc., 2020).
72. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. 2013–2015 (2015).
73. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
74. E. A. Fox, A. E. Wright, M. Fumagalli, F. G. Vieira, ngsLD: Evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* **35**, 3855–3856 (2019).
75. G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
76. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
77. P. Danecek, S. Schiffels, R. Durbin, Multiallelic calling model in bcftools (-m) (2014); https://samtools.github.io/bcftools/call-m.pdf.

78. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
79. N. Vijay, C. Park, J. Oh, S. Jin, E. Kern, H. W. Kim, J. Zhang, J.-K. Park, Population genomic analysis reveals contrasting demographic changes of two closely related dolphin species in the last glacial. *Mol. Biol. Evol.* **35**, 2026–2033 (2018).
80. M. Milanesi, S. Capomaccio, E. Vajana, L. Bomba, J. F. Garcia, P. Ajmone-Marsan, L. Colli, BITE: An R package for biodiversity analyses. bioRxiv, 181610 (2017). https://doi.org/10.1101/181610.
81. B. L. Taylor, S. J. Chivers, J. Larese, W. F. Perrin, Generation length and percent mature estimates for IUCN assessments of cetaceans (Administrative Report LJ-07-01). *National Marine Fisheries Service, Southwest Fisheries Science Center, NOAA, La Jolla, CA, USA* (2007).
82. A. Dornburg, M. C. Brandley, M. R. McGowen, T. J. Near, 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 (2012).
83. H.-S. Yim, Y. S. Cho, X. Guang, S. G. Kang, J.-Y. Jeong, S.-S. Cha, H.-M. Oh, J.-H. Lee, E. C. Yang, K. K. Kwon, Y. J. Kim, T. W. Kim, W. Kim, J. H. Jeon, S.-J. Kim, D. H. Choi, S. Jho, H.-M. Kim, J. Ko, H. Kim, Y.-A. Shin, H.-J. Jung, Y. Zheng, Z. Wang, Y. Chen, M. Chen, A. Jiang, E. Li, S. Zhang, H. Hou, T. H. Kim, L. Yu, S. Liu, K. Ahn, J. Cooper, S.-G. Park, C. P. Hong, W. Jin, H.-S. Kim, C. Park, K. Lee, S. Chun, P. A. Morin, S. J. O'Brien, H. Lee, J. Kimura, D. Y. Moon, A. Manica, J. Edwards, B. C. Kim, S. Kim, J. Wang, J. Bhak, H. S. Lee, J.-H. Lee, Minke whale genome and aquatic adaptation in cetaceans. *Nat. Genet.* **46**, 88–92 (2014).
84. R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing (Vienna, Austria, 2019); www.R-project.org/.
85. F. Racimo, S. Sankararaman, R. Nielsen, E. Huerta-Sánchez, Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359–371 (2015).
86. G. McVean, A genealogical interpretation of principal components analysis. *PLOS Genet.* **5**, e1000686 (2009).
87. T. Jombart, I. Ahmed, adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
88. E. Cuéllar-Rivas, M. C. Pustovrh-Ramos, The role of enamelysin (mmp-20) in tooth development: Systematic review. *Rev. Fac. Odontol. Univ. Antioq.* **27**, 154–176 (2015).

89. P. E. Rosel, B. A. Block, Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius*. *Mar. Biol.* **125**, 11–22 (1996).
90. S. A. Miller, D. D. Dykes, H. F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215–1215 (1988).
91. R. Sinha, G. Stanley, G. S. Gulati, C. Ezran, K. J. Travaglini, Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. bioRxiv (2017); www.biorxiv.org/content/biorxiv/early/2017/04/09/125724.full.pdf.
92. P. A. Morin, K. M. Parsons, F. I. Archer, M. C. Ávila-Arcos, L. G. Barrett-Lennard, L. Dalla Rosa, S. Duchêne, J. W. Durban, G. M. Ellis, S. H. Ferguson, J. K. Ford, M. J. Ford, C. Garilao, M. T. P. Gilbert, K. Kaschner, C. O. Matkin, S. D. Petersen, K. M. Robertson, I. N. Visser, P. R. Wade, S. Y. W. Ho, A. D. Foote, Geographic and temporal dynamics of a global radiation and diversification in the killer whale. *Mol. Ecol.* **24**, 3964–3979 (2015).
93. S. Andrews, Others, FastQC: A quality control tool for high throughput sequence data (2010).
94. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
95. K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B. J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody; Broad Institute Sequencing Platform and Whole Genome Assembly Team, K. C. Worley, C. L. Kovar, D. M. Muzny, R. A. Gibbs; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson; Genome Institute at Washington University, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
96. A. E. Moura, J. G. Kenny, R. Chaudhuri, M. A. Hughes, A. J. Welch, R. R. Reisinger, P. J. N. de Bruyn, M. E. Dahlheim, N. Hall, A. Rus Hoelzel, Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Mol. Ecol.* **23**, 5179–5192 (2014).
97. A. D. Foote, Y. Liu, G. W. C. Thomas, T. Vinař, J. Alföldi, J. Deng, S. Dugan, C. E. van Elk, M. E. Hunter, V. Joshi, Z. Khan, C. Kovar, S. L. Lee, K. Lindblad-Toh, A. Mancia, R. Nielsen, X. Qin, J.

Qu, B. J. Raney, N. Vijay, J. B. W. Wolf, M. W. Hahn, D. M. Muzny, K. C. Worley, M. T. P. Gilbert, R. A. Gibbs, Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275 (2015).

98. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
99. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
100. H. Wickham, Scales: Scale functions for visualization, *R package version 0. 4. 0* (2016); https://CRAN.R-project.org/package=scales.
101. E. Neuwirth, RColorBrewer: ColorBrewer palettes, R package version 1.1–2, *The R Foundation* (2014).
102. R. Nielsen, T. Korneliussen, A. Albrechtsen, Y. Li, J. Wang, SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLOS ONE* **7**, e37558 (2012).
103. B. J. Knaus, N. J. Grünwald, vcfr: A package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).
104. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
105. D. Maglott, J. Ostell, K. D. Pruitt, T. Tatusova, Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Res.* **39**, D52–D57 (2011).
106. The UniProt Consortium, UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
107. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

108. G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran, D. Lancet, The GeneCards suite: From gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* **54**, 1.30.1–1.30.33 (2016).
109. A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini, V. A. McKusick, Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–D517 (2005).