The dopamine receptor D_5 gene shows signs of independent erosion in toothed and baleen whales

Luís Q. Alves [1, 2], **Juliana Alves** [1, 2], **Rodrigo Ribeiro** [1, 2], **Raquel Ruivo** [Corresp., 1], **Filipe Castro** [Corresp. 1]

[1] CIIMAR-University of Porto, Matosinhos, Portugal

[2] FCUP- University of Porto, Porto, Portugal

Corresponding Authors: Raquel Ruivo, Filipe Castro
Email address: rruivo@ciimar.up.pt, filipe.castro@ciimar.up.pt

To compare gene *loci* considering a phylogenetic framework is a promising approach to uncover the genetic basis of human diseases. Imbalance of dopaminergic systems is suspected to underlie some emerging neurological disorders. The physiological functions of dopamine are transduced via G-protein-coupled receptors, including DRD_5 which displays a relatively higher affinity towards dopamine. Importantly, DRD_5 knockout mice are hypertense, a condition emerging from an increase in sympathetic tone. We investigated the evolution of DRD_5, a high affinity receptor for dopamine, in mammals. Surprisingly, among 124 investigated mammalian genomes, we found that Cetacea lineages (Mysticeti and Odontoceti) have independently lost this gene, as well as the burrowing Chrysochloris asiatica (Cape golden mole). We suggest that DRD_5 inactivation parallels hypoxia-induced adaptations, such as peripheral vasoconstriction required for deep-diving in Cetacea, in accordance with the convergent evolution of vasoconstrictor genes in hypoxia-exposed animals. Our findings indicate that Cetacea are natural knockouts for DRD_5 and might offer valuable insights into the mechanisms of some forms of vasoconstriction responses and hypertension in humans.

The dopamine receptor D_5 gene shows signs of independent erosion in Toothed and Baleen whales

Luís Q. Alves[1,2], Juliana Alves[1,2], Rodrigo Ribeiro[1,2], Raquel Ruivo[1], L. Filipe C. Castro[1,2]

[1] CIIMAR-University of Porto, Matosinhos, Portugal
[2] FCUP-University of Porto, Porto, Portugal

Corresponding Authors:
Filipe Castro and Raquel Ruivo
Avenida General Norton de Matos, S/N, Matosinhos, 4450-208, Matosinhos, Portugal
Email address: rruivo@ciimar.up.pt; filipe.castro@ciimar.up.pt

Abstract

To compare gene *loci* considering a phylogenetic framework is a promising approach to uncover the genetic basis of human diseases. Imbalance of dopaminergic systems is suspected to underlie some emerging neurological disorders. The physiological functions of dopamine are transduced via G-protein-coupled receptors, including DRD_5 which displays a relatively higher affinity towards dopamine. Importantly, DRD_5 knockout mice are hypertense, a condition emerging from an increase in sympathetic tone. We investigated the evolution of DRD_5, a high affinity receptor for dopamine, in mammals. Surprisingly, among 124 investigated mammalian genomes, we found that Cetacea lineages (Mysticeti and Odontoceti) have independently lost this gene, as well as the burrowing *Chrysochloris asiatica (*Cape golden mole). We suggest that DRD_5 inactivation parallels hypoxia-induced adaptations, such as peripheral vasoconstriction required for deep-diving in Cetacea, in accordance with the convergent evolution of vasoconstrictor genes in hypoxia-exposed animals. Our findings indicate that Cetacea are natural knockouts for DRD_5 and might offer valuable insights into the mechanisms of some forms of vasoconstriction responses and hypertension in humans.

Introduction

Dopamine is a neurotransmitter essential for brain function, regulating various physiological processes including locomotion, cognition, and neuroendocrine functions (Hollon et al., 2002; Ott & Nieder, 2018). Dopamine molecular actions are transduced *via* a specific group of G-protein coupled receptors entailing two major classes: DRD_1-like and DRD_2-like receptors (Beaulieu & Gainetdinov, 2011; Opazo et al., 2018). While DRD_1-like receptors stimulate cAMP production postsynaptically, DRD_2-like receptors inhibit cAMP production both pre and postsynaptically (Beaulieu & Gainetdinov, 2011). The genomic structure of the underlying genes is also distinct, with DRD_1-like receptors yielding single exon coding regions (Beaulieu & Gainetdinov, 2011). $DRD_{1\text{-}2}$ receptor classes have diversified in vertebrate evolution most likely as a result of genome duplications (Opazo et al., 2018). Interestingly, agonist and antagonist amino acid site conservation suggests evolutionary stasis of dopaminergic pathways (Opazo et al., 2018). Among DRD_1-like receptors, the DRD_5 subtype displays distinctive features, namely a relatively higher affinity towards dopamine, a putative agonist-independent activity and low level, yet widespread, brain expression (Beaulieu & Gainetdinov, 2011; Ciliax et al., 2000; Sunahara et al., 1991; Tiberi & Caron, 1994). Nonetheless, the DRD_5 seems to display distinct regional and cellular distribution patterns in the brain, when compared to the DRD_1 and DRD_2 subtypes, with protein enrichment detected in the cerebral cortex, hippocampus and basal ganglia (Ariano et al., 1997; Ciliax et al., 2000). Peripheral expression has also been found in the adrenals (Dahmer & Senogles, 1996), kidney (Sanada et al., 2000) and gastrointestinal tract (Mezey et al., 1996). Despite the association with schizophrenia (Muir et al., 2001), attention-deficit/hyperactive disorder (Daly et al., 1999) and substance abuse (Vanyukov et al., 1998), gene targeting studies revealed that DRD_5 knock-out mice develop hypertension, showing increased blood pressure from 3 months of age (Hollon et al., 2002). This hypertensive phenotype appears to result from a central nervous system defect, leading to an increase in sympathetic tone and, consequently, vasoconstriction (Hollon et al., 2002). Besides neuronal impairment, DRD_5 disruption was also suggested to increase the expression of the prohypertensive Angiotensin II Type 1 Receptor (AT_1R), involved in renal salt balance, blood pressure and vasoconstriction (Li et al., 2008). In fact, renal dopamine was suggested to promote salt excretion, counter-regulating AT_1R and lowering hypertensive states (Hong et al., 2017; Li et al., 2008; Zeng et al., 2005).

Whole genome sequencing has greatly expanded our capacity to comprehend evolutionary history, the role of adaptation or the basis for phenotype differences across the tree of life. Multi-genome comparisons have also been powerful to recognize the molecular basis of human diseases, a field named as phylomedicine (Emerling et al., 2017; Kumar et al., 2011; Miller & Kumar, 2001; Springer & Murphy, 2007). Here we investigate the evolution of DRD_5 in mammalian species. By analyzing 124 genomes covering 16 orders, we show that independent coding-debilitating mutations occurred in the ancestors of Mysticeti, of *Physeter macrocephalus* (sperm whale) and the remaining Odontoceti, strongly suggesting that DRD_5 is non-coding in Cetacea. Reductive episodes have been widely documented across the tree of life contributing to organismal divergence and physiological and morphological adaptation to environmental cues (Albalat & Canestro, 2016; Braun, 2003; Jeffery, 2009; Olson, 1999). In agreement, gene loss mechanisms seem pervasive in lineages that endured drastic habitat transitions in the course of evolution, such as Cetacea, entailing niche-specific adaptations (Huelsmann et al., 2019; Lachner et al., 2017; Lopes-Marques et al., 2019b; McGowen et al., 2014; Nery et al., 2014; Sharma et al., 2018; Strasser et al., 2015). Thus, our findings suggest that these species are natural KOs for this dopamine receptor and might offer valuable insights into the mechanisms of some forms of essential hypertension.

Materials & Methods

To manually infer the coding status of DRD_5 genes, the following strategy was used: first, gene orthology was assessed *via* synteny analysis, to clarify cases where gene annotation was not found, as well as to define genomic regions to be posteriorly collected for gene annotation (see Supplementary Figure 1). Next, we performed the manual annotation of the open reading frame (ORF), and the correspondent sequence screened for abolishing mutations. At least one mutation was posteriorly validated using raw genomic sequencing reads available at the NCBI (National Center of Biotechnology Information) Sequence Read Archive (SRA) database. Finally, d_N/d_S analyses were also conducted to further investigate the inactive status of DRD_5 in Cetacea (see below).

Synteny analysis

To build the synteny maps for the DRD_5 gene *locus* in Cetacea and *Hippopotamus amphibius* (common hippopotamus) several annotated Cetacea genome assemblies were inspected and scrutinised using the NCBI browser, namely *Orcinus orca* (killer whale; GCF_000331955.2), *Lagenorhynchus obliquidens* (Pacific white-sided dolphin; GCF_003676395.1), *Tursiops truncatus* (common bottlenose dolphin; GCF_001922835.1), *Delphinapterus leucas* (beluga whale; GCF_002288925.1), *Neophocaena asiaeorientalis asiaeorientalis* (Yangtze finless porpoise; GCF_003031525.1), *Lipotes vexillifer* (Yangtze River dolphin; GCF_000442215.1), *Physeter macrocephalus* (sperm whale; GCF_002837175.1) and *Balaenoptera acutorostrata scammoni* (minke whale; GCF_000493695.1). *Bos taurus* (cattle; GCF_002263795.1), a fully terrestrial relative of extant cetaceans, was used as reference. Next, (1) in genome assemblies with annotated DRD_5, the following procedure was used: five protein-coding genes, upstream and downstream of the DRD_5 gene, and from the same strand, were collected; (2) if DRD_5 gene annotations were not present, the genomic *locus* was retrieved using *Bos taurus* (cattle) DRD_5 flanking genes as reference. The selected anchoring genes to search upstream and downstream DRD_5 neighbouring genes were CPEB2 and OTOP1, respectively. Regarding *Hippopotamus amphibius* (common hippopotamus), the synteny map was built via blast searches against the assembled, fragmented and unannotated genome of the same species, available at NCBI (GCA_002995585.1). *Bos taurus* (cattle) DRD_5 flanking genes were used as reference; using the discontiguous megablast task from blastn, the best blast hit (highest alignment identity and query coverage) was retrieved and the coordinates of the alignment in the target genome carefully inspected. The *Hippopotamus amphibius* (common hippopotamus) synteny map was then built by sorting the genes according to the subject alignment coordinates within genes aligning at the same genomic scaffold.

Sequence retrieval and gene annotation

The NCBI 'low-quality protein' (LQ) tag marks RefSeq sequences that have been automatically modified relative to the corresponding genome sequence to correct for possible open reading frame (ORF) protein-altering indels or mismatches, assuming that these arise from sequencing errors or genome assembly artefacts. Yet, in several cases these frameshift and nonsense codons correspond to authentic biological modifications leading to shifts or truncations

of the predicted ORF, culminating into an abnormal protein amino acid constitution. For this reason, to clarify the functional status of DRD_5 in species presenting LQ tag for this gene, the corresponding genomic regions were directly collected from NCBI.

Concerning species with no annotated DRD_5 genes two scenarios were identified: (1) annotated genomes excluding DRD_5 gene annotation and (2) unannotated genomes. Regarding the first case (i.e. the cetaceans *Lipotes vexillifer,* the Yangtze River dolphin and *Balaenoptera acutorostrata scammoni*, the minke whale, as well as other non-cetacean mammals), the DRD_5 genomic *locus* was retrieved using, as reference, annotated DRD_5 flanking genes from closely related species (for cetaceans, the terrestrial extant sister clade *Bos taurus* (cattle), regarding non-cetacean mammals see Supplementary Table 1). If the genomic *locus* exhibited severe genomic fragmentation (presence of Ns), thus, hindering the retrieval using neighbouring genes, blastn searches were conducted against the Whole Shotgun Contigs of the corresponding species *via* discontiguous megablast task, using as query the DRD_5 coding sequence of the same reference species (Supplementary Table 1). The genomic sequence corresponding to the blast hit with the highest alignment identity and query coverage was selected.

For species without annotated genomes (i.e. *Balaenoptera bonaerensis* (Antarctic minke whale)*, Eschrichtius robustus* (grey whale)*, Balaena mysticetus* (bowhead whale)*, Sousa chinensis* (Indo-pacific humpbacked dolphin)*,* as well as *Hippopotamus amphibius* (hippopotamus)) genomic sequences were retrieved through blastn searches in the corresponding genome assembly using the *Bos taurus* (cattle) DRD_5 coding sequence as query. For each species, the best genomic scaffold corresponded to the blast hit with the highest query coverage and identity value. Due to the presence of a fragmented genomic region in the DRD_5 gene annotation of *T. truncatus* (common bottlenose dolphin), the same blast search procedure was carried out for this species, using as target the Whole Genome Shotgun contig dataset of the same species.

Collected genomic sequences were further imported into Geneious Prime 2019 (www.geneious.com) and the DRD_5 gene coding sequences manually annotated for each species. Briefly, using the built-in map to reference tool with the highest sensibility parameter selected, the reference single-exon DRD_5 gene, 3' and 5' UTR flanked, was mapped against the corresponding genomic sequence of the in-study species. Aligned regions were further carefully screened for ORF abolishing mutations including frameshift mutations and in-frame premature

stop codons. For Cetacea and *Hippopotamus amphibius* (common hippopotamus) DRD_5 gene annotation, *Bos taurus* (cattle) DRD_5 was selected as reference. Regarding non-cetacean mammals DRD_5 annotation, different references were chosen according to the phylogenetic relationships between reference and test species (Supplementary Table 1). Finally, the identified mutations were next validated using raw sequencing reads retrieved from at least two independent genomic NCBI sequence read archive (SRA) projects (when available). Briefly, blastn searches were conducted against the selected SRA projects (Supplementary Materials 1, 2 and 3) using as query the nucleotide sequence containing the mutation. Blast hits were imported into Geneious Prime 2019 and mapped against the manually annotated sequences, using the built-in map to reference tool, to confirm the presence of the identified mutation.

Phylogenetic and d_N/d_S analyses

The predicted cetacean and *Hippopotamus amphibius* (common hippopotamus) DRD_5 coding sequences, as well as the DRD_5 coding sequences of *Homo sapiens* (human) and *Bos taurus* (cattle) available at NCBI were translation aligned in Geneious Prime 2019 using the Blosum62 substitution matrix. The alignment (1440 bp) was manually inspected and predicted DRD_5 coding sequences involved in alignment suffered prior inspection, with frameshift insertions and deletions being omitted and stop codons recorded as missing. The alignment was posteriorly exported for phylogenetic analysis. A Maximum likelihood phylogenetic tree, performed in PhyML3.0 server (Guindon et al., 2010), was produced with the best sequence evolutionary model being determined using the built-in smart model selection (HKY85 +G) (Lefort et al., 2017). Node support was inferred based on 1000 bootstrap replicates. The resulting phylogenetic tree newick file is available as supplementary data (Supplementary File 1) and was subsequently imported for visualisation in FigTree (Supplementary Material 4) (Rambaut & Drummond, 2012).

PAML 4.6 (Yang, 2007) was used to estimate the ratio (ω) of the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S) using the produced phylogenetic tree (Supplementary Material 4). Codeml with the branch model was implemented to estimate d_N/d_S ratios (ω) for seven different branch categories: Mysticeti, Odontoceti, Functional, a category corresponding to the common cetacean branch, other concerning the stem Mysticeti

196 branch, an additional one regarding the stem Odontoceti branch (excluding *Physeter*
197 *macrocephalus,* the sperm whale) and finally, a category corresponding to the *Physeter*
198 *macrocephalus* (sperm whale) ancestor branch (Table 1). Mysticeti and Odontoceti branch
199 categories comprised all predicted Mysticeti and Odontoceti DRD_5 sequences, respectively. The
200 Functional branch category comprised *Hippopotamus amphibius* (common hippopotamus), *Bos*
201 *taurus* (cattle) and *Homo sapiens* (human) DRD_5 coding sequences. For each of the seven
202 different branch categories, likelihood ratio tests (LRT) (following χ^2 distribution) were
203 conducted to compare the estimated ratio (ω) of the nonsynonymous substitution rate (d_N) to the
204 synonymous substitution rate (d_S) determined by PAML (the alternative hypothesis), with the
205 expected ratio (ω) of the nonsynonymous substitution rate (d_N) to the synonymous substitution
206 rate (d_S) according to a neutral model of evolution ($\omega = 1$) (the null hypothesis). All PAML
207 analyses were run with the CodonFreq set to 3 and codon sites with ambiguous data (including
208 gaps and missing data) were included in the analyses.

209

210 ## Results

211 To examine the annotation tags and distribution of the DRD_5 gene across mammals, 119
212 annotated mammalian genomes available at NCBI were scrutinized for the presence of DRD_5
213 gene annotation and each respective protein product description screened for the 'low-quality
214 protein' (LQ) tag. This examination resulted in 10 species presenting the DRD_5 LQ tag,
215 including *Ovis aries* (sheep), *Phascolarctos cinereus* (koala), *Bison bison bison* (plains bison),
216 *Myotis davidii* (vesper bat), *Ochotona princeps* (American pika) and 5 cetacean species. The
217 latter included *Lagenorhynchus obliquidens* (Pacific white-sided dolphin), *N. a. asiaeorientalis*
218 (Yangtze finless porpoise), *D. leucas* (beluga whale), *Physeter macrocephalus* (sperm whale)
219 *and Orcinus orca* (killer whale). Each genomic sequence corresponding to the DRD_5 LQ
220 annotations was examined and the coding sequence (CDS) manually predicted (Lopes-Marques
221 et al., 2017). Given the prominence of DRD_5 LQ annotations in Cetacea we scrutinized other
222 cetacean species with available, but unannotated genomes, *Balaenoptera bonaerensis* (Antarctic
223 minke whale), *Eschrichtius robustus* (gray whale), *Balaena mysticetus* (bowhead whale), *S.*
224 *chinensis* (Indo-Pacific humpback dolphin), or with annotated genomes lacking DRD_5
225 annotations: *Lipotes vexillifer* (Yangtze River dolphin) and *Balaenoptera acutorostrata*

scammoni (minke whale) (Supplementary Figure 1). Additionally, *T. truncatus* (bottlenose dolphin), presenting a seemingly intact DRD_5 gene annotation, without the 'low-quality protein' tag (LQ), as well as *Hippopotamus amphibius* (common hippopotamus) predicted DRD_5 coding sequence, representing the closest extant lineage of Cetacea, were equally inspected. Other mammals with annotated genome without DRD_5 gene annotation were also scrutinised, namely: *Microcebus murinus* (gray mouse lemur), *Jaculus jaculus* (lesser Egyptian jerboa), *Chrysochloris asiatica* (Cape golden mole), *Erinaceus europaeus* (western European hedgehog), *Elephantulus edwardii* (Cape elephant shrew) and *Condylura cristata* (star-nosed mole). In total, 124 mammalian species were inspected and an in-depth description regarding analysed species list and genomic sequences accession numbers are available at Supplementary Table 2. The Supplementary Figure 2 presents a multiple translation alignment of the NCBI non-LQ tagged mammalian DRD_5 orthologous sequences. The alignment also includes the predicted *Hippopotamus amphibius* (hippopotamus) DRD_5 sequence and excludes *T. truncatus* (common bottlenose dolphin) DRD_5 sequence, afterwards demonstrated to contain inactivating mutations (see below). The examined sequences exhibit a substantial degree of conservation (average pairwise identity of over 80%), with minor variation in the expected protein size. The protein sequence conservation is particularly noticeable at the c-terminus (Supplementary Figure 2).

ORF-disrupting mutations of DRD_5 in Cetacea

For cetacean species with annotated genomes, we started by examining the DRD_5 gene *locus*, including neighbouring genes, to verify and elucidate the orthology of the annotated and non-annotated genes and outline the genomic regions to be inspected (Supplementary Figure 1). All analysed *loci* were found to be conserved, including in both *Lipotes vexillifer* (Yangtze River dolphin) and *Balaenoptera acutorostrata scammoni* (minke whale), which lacked previous DRD_5 gene annotations (Supplementary figure 1). Subsequent manual annotation of all collected cetacean genomic sequences revealed DRD_5 gene erosion across all analysed species, except in *Balaenoptera acutorostrata scammoni* (minke whale), for which the DRD_5 coding status could not be accessed due to fragmentation of the 5' end of the respective genomic region (presence of sequencing gaps (Ns)). In detail, a conserved 2 nucleotide deletion was detected for the full set of Odontoceti examined species, except for *Physeter macrocephalus* (sperm whale) that presented a premature stop codon near the middle of the gene and a single nucleotide insertion close to the

end of the gene (Figure 1). The 2-nucleotide deletion alters the reading frame, leading to a drastic change in downstream amino acid composition. Additionally, non-conserved mutations were found in *Lipotes vexillifer* (Yangtze River dolphin), which presented 2 premature stop codons, *N. a. asiaeorientalis* (Yangtze finless porpoise) that presented a premature stop codon close to the middle of the ORF, and *D. leucas* (beluga whale) with a single nucleotide deletion near the 5' end of the gene (Figure 1). *D. leucas* (beluga whale) presumed DRD_5 sequence presented another noticeable feature. A massive and abrupt alignment identity decrease observed when aligning *Bos taurus* (cattle) DRD_5 gene against the genomic target region of *D. leucas* (beluga whale). The alignment identity drop was noted approximately in the middle of the complete alignment length for this species, suggesting that the DRD_5 gene sequence is interrupted, and further supporting pseudogenization in this species. Regarding Odontoceti, at least one ORF-abolishing mutation was validated using available genomic Sequence Read Archives (SRAs) experiments for all studied species, excluding *S. chinensis* (Indo-pacific humpbacked dolphin) and *Lipotes vexillifer* (Yangtze River dolphin) for which no genomic sequencing runs were available at the NCBI SRA database (Supplementary Material 1).

Regarding the Mysticeti suborder, a conserved single nucleotide deletion was detected in all species except *Balaenoptera acutorostrata scammoni* (minke whale) (Figure 1). A non-conserved 2 nucleotide insertion was found also found in *Balaenoptera bonaerensis* (Antartic minke whale) and an insertion of one nucleotide was detected at *Eschrichtius robustus* (grey whale) near the 5' end of the DRD_5 sequence (Figure 1). Again, at least one ORF-abolishing mutation was validated using genomic SRA experiments for all analysed species (Supplementary Material 1). Importantly, no conserved mutations were detected between most Odontoceti, *Physeter macrocephalus* (sperm whale, Odontoceti) and Mysticeti lineages, suggesting that DRD_5 three pseudogenization events occurred independently after their evolutionary divergence (Figure 1). To increase the robustness of our analysis, we further scrutinized the genome of the extant sister clade of the Cetacea, the Hippopotamidae and were able to predict a fully coding sequence for DRD_5 in *Hippopotamus amphibius* (common hippopotamus), supporting the loss of DRD_5 after Cetacea diversification.

d_N/d_S analyses support independent inactivation events of DRD_5 in Cetacea lineages

To further strengthen the mutational analyses, we next carried out a d_N/d_S approach. The d_N/d_S analyses for species with functional DRD_5 gene (Functional branches category) rejected the null hypothesis ($\omega = 0.09$, $p = 0.030$, $< .05$), confirming that, as expected, these have evolved under purifying selection ($\omega < 1$), favoring the conservation of DRD_5 in the tested lineages (Table 1). In contrast, concerning Mysticeti and Odontoceti branch categories, the performed analyses failed to reject the null hypothesis ($\omega = 0.386$, $p = 0.692$ for Odontoceti branches category and $\omega = 0.460$, $p = 0.291$ for Mysticeti branches category, respectively) suggesting that DRD_5 evolved neutrally in both lineages, following gene inactivation events (Table 1). Moreover, evidences of purifying selection were found for the common Cetacean branch category ($\omega = 0.135$, $p = 0.001$), suggesting the existence of a functional DRD_5 in the common ancestor to all cetaceans (Table 1). Purifying selection was also detected for the Odontoceti stem branch category (excluding *Physeter macrocephalus*, the sperm whale) ($\omega = 0.213$, $p = 0.045$). In contrast the *Physeter macrocephalus* (sperm whale) ancestor branch category, does not present a d_N/d_S ratio (ω) statistically significantly different from 1 ($\omega = 0.320$, $p = 0.460$) (Table 1). Together with the mutational evidences (Figure 1), these results suggest two independent inactivation events within the Odontoceti. Regarding the Mysticeti stem branch category, selection analysis suggested neutral evolution ($\omega = 1.174$, $p = 0.905$) (Table 1). However, mutational evidence supports a conserved inactivation event in the common ancestor of Mysticeti. Taken together, our results support at least two independent DRD_5 inactivation events within Cetacea.

Other mammalian species displaying DRD_5 LQ tags have a coding gene

Initial analysis revealed the presence of at least one ORF-abolishing mutation in *Ovis aries* (sheep), *Phascolarctos cinereus* (koala), *Bison bison bison* (plains bison) and *Ochotona princeps* (American pika). These are in some cases suggestive of gene inactivation and not sequencing artefacts (e.g. Emerling et al., 2017; Lopes-Marques et al., 2017). Manual annotation, including SRA validation, unveiled sequencing reads supporting the absence of disruptive ORF mutations, rebutting each inactivation mutation and suggesting that DRD_5 is, in fact, coding in these species (Supplementary Material 2). Regarding *Myotis davidii* (vesper bat), the fragmentation of the genomic region (Ns) flanked by upstream and downstream DRD_5 neighbouring genes impeded

us to infer the DRD_5 coding status in this species. Interestingly, regarding *Ochotona princeps* (American pika) mutational SRA validation, two scenarios were observed: approximately 50% of aligned reads supported the presence of a premature stop codon in the DRD_5 gene of this species, with the remaining set of aligning reads supporting the absence of a premature stop codon in the same species. This suggest that a polymorphic loss event might have occurred in this species, with non-functional and functional alleles currently segregating in the correspondent population.

***Chrysochloris asiatica* presents a non-functional DRD_5 gene**

Next, we examined other mammalian species with annotated genome yet lacking DRD_5 gene annotations. Results were inconclusive regarding the coding status of *Microcebus murinus* (gray mouse lemur), *Condylura cristata* (star-nosed mole), *J. jaculus* (lesser Egyptian jerboa) and *Erinaceus europaeus* (western European hedgehog) DRD_5, due to the fragmentation (Ns) of the genomic region flanked by the upstream and downstream DRD_5 neighbouring genes, and the unavailability of whole genome shotgun contigs spanning our target gene. For *Elephantulus edwardii* (Cape elephant shrew) we were able to deduce a fully DRD_5 coding sequence. Curiously, *Chrysochloris asiatica* (Cape golden mole) presented a single nucleotide insertion in the 5' end of the gene (validated by genomic SRA, see Supplementary Material 3), suggesting that DRD_5 might be pseudogenized in this species.

Discussion

The rise of large-scale genomic sequencing projects has emphasized the role of gene loss as a potent driver of evolutionary change: underlying phenotypic adaptations or neutral regressions in response to specific environmental cues and niches (Albalat & Canestro, 2016; Braun, 2003; Jebb & Hiller, 2018; Jeffery, 2009; Olson, 1999; Sadier et al., 2018; Sharma et al., 2018; Somorjai et al., 2018). By comparing 124 mammalian genomes, we document three independent erosion events of a dopamine receptor, DRD_5, in Cetacea lineages. Although the current analysis is highly dependent of the quality of genome sequencing projects and their assembly, the conserved mutational profile of DRD_5 gene inactivation within Cetacea lineages further strengthens our findings.

Dopamine, a neurotransmitter and signalling molecule, is involved in distinct functions both in the central nervous system and peripheral tissues: including movement, feeding, sleep, reward, learning and memory as well as in the regulation of olfaction, hormone pathways, renal functions, immunity, sympathetic regulation and cardiovascular functions, respectively (Beaulieu & Gainetdinov, 2011). More specifically, the disruption of DRD_5-dependent pathways in rodents was shown to increase blood pressure and sympathetic tone, promoting vasoconstriction, thus yielding a hypertensive phenotype (Hollon et al., 2002; Li et al., 2008).

The observed gene loss distribution, and predicted phenotypic outcome, is consistent with the peripheral vasoconstriction mechanism described in Cetacea, suggested to counterbalance deep-diving induced hypoxia (Ramirez et al., 2007; Tian et al., 2016). In fact, Cetacea have developed a number of physiological adaptations to offset hypoxia: notably, increased blood volume and oxygen-transport protein levels (hemoglobin, neuroglobin and myoglobin), allowing oxygen stores in blood, muscle tissues, and brain, reduced heart rate or bradycardia, apnea and peripheral vasoconstriction (Nery et al., 2013; Panneton, 2013; Ramirez et al., 2007; Tian et al., 2016). Peripheral vasoconstriction allows the regional compartmentalization of blood supplies, reducing blood flow in more hypoxia-tolerant tissues, such as skin, muscle, spleen or kidney, while maintaining arterial blood flow to the central nervous system and heart (Panneton, 2013). Moreover, the deviation of muscle blood supply reduces lactate accumulation (Panneton, 2013). Thus, in Cetacea, DRD_5 loss could contribute for the peripheral vasoconstriction requirements of diving. Also, by shifting renal salt balance, DRD_5 could also play a role in the maintenance of an adequate blood volume and pressure (Bie, 2009; Hong et al., 2017; Li et al., 2008; Zeng et al., 2005).

The predicted vasoconstriction phenotype is in agreement with a previous work reporting episodes of adaptive evolution (positive selection) in genes related with hypoxia tolerance in Cetacea, including genes involved in oxygen transport and regulation of vasoconstriction (Tian et al., 2016). In addition, by expanding their analysis to other non-aquatic hypoxic environments, such as underground tunnels, they uncovered convergent evolution scenarios in species adapted to diving and burrowing (Tian et al., 2016). A similar convergence is observed in the present work. In fact, in the mole *Chrysochloris asiatica* (Cape golden mole), DRD_5 was also predicted non-functional. Although this was the single DRD_5 loss example found outside Cetacea, one cannot discard the putative contribution of alternative molecular events (i.e. post-translational

mechanisms) towards trait loss (Sadier et al., 2018), or event distinct physiological adaptations to overcome oxygen deprivation.

Besides diving physiology, DRD_5-dependent sympathetic tone alterations could also contribute to the idiosyncratic sleep behavior observed in Cetacea (Huelsmann et al., 2019; Lopes-Marques et al., 2019b; Lyamin et al., 2008). Several physiological adjustments occur during sleep, encompassing thermoregulation, as well as endocrine, immune, pulmonary and cardiovascular functions (Giglio et al., 2007). In most mammals, sleep states lead to a decrease of the sympathetic tone, inducing vasodilation and decreasing blood pressure (Giglio et al., 2007). Thus, DRD_5 loss could prevent sympathetic tone decrease in resting states paralleling the unihemispheric sleeping behavior and long-term vigilance observed in Cetacea.

Conclusions

Overall, our findings provide evidence for natural occurring KO for DRD_5. Besides highlighting a molecular signature for vasoconstriction and blood pressure regulation in Cetacea, naturally occurring DRD_5 KO could also provide useful frameworks to gain insight into hypertension and heart failure-induced peripheral vasoconstriction responses in humans (Triposkiadis et al., 2009; Wang et al., 2008).

Acknowledgements

We acknowledge the various genome consortiums for sequencing and assembling the genomes.

References

Albalat R, Canestro C. 2016. Evolution by gene loss. *Nature Reviews Genetics* 17:379-391. 10.1038/nrg.2016.39

Ariano MA, Wang J, Noblett KL, Larson ER, Sibley DR. 1997. Cellular distribution of the rat D1B receptor in central nervous system using anti-receptor antisera. *Brain Research* 746:141-150. https://doi.org/10.1016/S0006-8993(96)01219-X

Beaulieu J-M, Gainetdinov RR. 2011. The Physiology, Signaling, and Pharmacology of Dopamine Receptors. *Pharmacological Reviews* 63:182. 10.1124/pr.110.002642

Bie P. 2009. Blood volume, blood pressure and total body sodium: internal signalling and output control. *Acta Physiologica* 195:187-196. 10.1111/j.1748-1716.2008.01932.x

Braun EL. 2003. Innovation from reduction: gene loss, domain loss and sequence divergence in genome evolution. *Applied Bioinformatics* 2:13-34.

Ciliax BJ, Nash N, Heilman C, Sunahara R, Hartney A, Tiberi M, Rye DB, Caron MG, Niznik HB, Levey AI. 2000. Dopamine D5 receptor immunolocalization in rat and monkey brain. *Synapse* 37:125-145. 10.1002/1098-2396(200008)37:2<125::AID-SYN7>3.0.CO;2-7

Dahmer MK, Senogles SE. 1996. Dopaminergic Inhibition of Catecholamine Secretion from Chromaffin Cells: Evidence that Inhibition Is Mediated by D4 and D5 Dopamine Receptors. *Journal of Neurochemistry* 66:222-232. 10.1046/j.1471-4159.1996.66010222.x

Daly G, Hawi Z, Fitzgerald M, Gill M. 1999. Mapping susceptibility loci in attention deficit hyperactivity disorder: preferential transmission of parental alleles at DAT1, DBH and DRD5 to affected children. *Molecular Psychiatry* 4:192. 10.1038/sj.mp.4000510

Emerling CA, Widjaja AD, Nguyen NN, Springer MS. 2017. Their loss is our gain: regressive evolution in vertebrates provides genomic models for uncovering human disease loci. *Journal of Medical Genetics* 54:787. 10.1136/jmedgenet-2017-104837

Giglio P, Lane JT, Barkoukis TJ, Dumitru I. 2007. Sleep Physiology. In: Barkoukis TJ, and Avidan AY, eds. *Review of Sleep Medicine (Second Edition)*. Philadelphia: Butterworth-Heinemann, 29-41.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59:307-321. 10.1093/sysbio/syq010

Hollon TR, Bek MJ, Lachowicz JE, Ariano MA, Mezey E, Ramachandran R, Wersinger SR, Soares-da-Silva P, Liu ZF, Grinberg A, Drago J, Young WS, Westphal H, Jose PA, Sibley DR. 2002. Mice Lacking D5 Dopamine Receptors Have Increased Sympathetic Tone and Are Hypertensive. *The Journal of Neuroscience* 22:10801. 10.1523/JNEUROSCI.22-24-10801.2002

Hong K, Li M, Nourian Z, Meininger GA, Hill MA. 2017. Angiotensin II Type 1 Receptor Mechanoactivation Involves RGS5 (Regulator of G Protein Signaling 5) in Skeletal Muscle Arteries: Impaired Trafficking of RGS5 in Hypertension. *Hypertension (Dallas, Tex : 1979)* 70:1264-1272. 10.1161/HYPERTENSIONAHA.117.09757

Huelsmann M, Hecker N, Springer MS, Gatesy J, Sharma V, Hiller M. 2019. Genes lost during the transition from land to water in cetaceans highlight genomic changes involved in aquatic adaptations. *biorxivorg*. 10.1101/521617

Jebb D, Hiller M. 2018. Recurrent loss of HMGCS2 shows that ketogenesis is not essential for the evolution of large mammalian brains. *eLife* 7. 10.7554/eLife.38906

Jeffery WR. 2009. Regressive evolution in Astyanax cavefish. *Annual Review of Genetics* 43:25-47. 10.1146/annurev-genet-102108-134216

Kumar S, Dudley JT, Filipski A, Liu L. 2011. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends in Genetics* 27:377-386. 10.1016/j.tig.2011.06.004

Lachner J, Mlitz V, Tschachler E, Eckhart L. 2017. Epidermal cornification is preceded by the expression of a keratinocyte-specific set of pyroptosis-related genes. *Scientific Reports* 7:17446. 10.1038/s41598-017-17782-4

Lefort V, Longueville J-E, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution* 34:2422-2424. 10.1093/molbev/msx149

Li H, Armando I, Yu P, Escano C, Mueller SC, Asico L, Pascua A, Lu Q, Wang X, Villar VAM, Jones JE, Wang Z, Periasamy A, Lau Y-S, Soares-da-Silva P, Creswell K, Guillemette G, Sibley DR, Eisner G, Felder RA, Jose PA. 2008. Dopamine 5 receptor mediates Ang II type 1 receptor degradation via a ubiquitin-proteasome pathway in mice and human cells. *The Journal of Clinical Investigation* 118:2180-2189. 10.1172/JCI33637

Lopes-Marques M, Machado AM, Alves QL, Fonseca MM, Barbosa S, Sinding M-HS, Rasmussen MH, Iversen MR, Bertelsen MF, Campos PF, da Fonseca R, Ruivo R, Castro LFC. 2019a. Complete inactivation of sebum-producing genes parallels the loss of sebaceous glands in Cetacea (in press). *Molecular Biology and Evolution*.

Lopes-Marques M, Machado AM, Barbosa S, Fonseca MM, Ruivo R, Castro LFC. 2018. Cetacea are natural knockouts for IL20. *Immunogenetics*. 10.1007/s00251-018-1071-5

Lopes-Marques M, Ruivo R, Alves QL, Sousa N, Machado MA, Castro FL. 2019b. The Singularity of Cetacea Behavior Parallels the Complete Inactivation of Melatonin Gene Modules. *Genes* 10. 10.3390/genes10020121

Lopes-Marques M, Ruivo R, Fonseca E, Teixeira A, Castro LFC. 2017. Unusual loss of chymosin in mammalian lineages parallels neo-natal immune transfer strategies. *Molecular Phylogenetics and Evolution* 116:78-86. https://doi.org/10.1016/j.ympev.2017.08.014

Lyamin OI, Manger PR, Ridgway SH, Mukhametov LM, Siegel JM. 2008. Cetacean sleep: An unusual form of mammalian sleep. *Neuroscience & Biobehavioral Reviews* 32:1451-1484. 10.1016/j.neubiorev.2008.05.023

McGowen MR, Gatesy J, Wildman DE. 2014. Molecular evolution tracks macroevolutionary transitions in Cetacea. *Trends in Ecology & Evolution* 29:336-346. https://doi.org/10.1016/j.tree.2014.04.001

Mezey E, Eisenhofer G, Harta G, Hansson S, Gould L, Hunyady B, Hoffman BJ. 1996. A novel nonneuronal catecholaminergic system: exocrine pancreas synthesizes and releases dopamine. *Proceedings of the National Academy of Sciences* 93:10377. 10.1073/pnas.93.19.10377

Miller MP, Kumar S. 2001. Understanding human disease mutations through the use of interspecific genetic variation. *Human Molecular Genetics* 10:2319-2328. 10.1093/hmg/10.21.2319

Muir WJ, Thomson ML, McKeon P, Mynett-Johnson L, Whitton C, Evans KL, Porteous DJ, Blackwood DHR. 2001. Markers close to the dopamine D5 receptor gene (DRD5) show significant association with schizophrenia but not bipolar disorder. *American Journal of Medical Genetics* 105:152-158. 10.1002/1096-8628(2001)9999:9999<::AID-AJMG1163>3.0.CO;2-2

Nery MF, Arroyo JI, Opazo JC. 2013. Accelerated Evolutionary Rate of the Myoglobin Gene in Long-Diving Whales. *Journal of Molecular Evolution* 76:380-387. 10.1007/s00239-013-9572-1

Nery MF, Arroyo JI, Opazo JC. 2014. Increased rate of hair keratin gene loss in the cetacean lineage. *BMC Genomics* 15:869. 10.1186/1471-2164-15-869

Olson MV. 1999. When less is more: gene loss as an engine of evolutionary change. *American Journal of Human Genetics* 64:18-23. 10.1086/302219

Opazo JC, Zavala K, Miranda-Rottmann S, Araya R. 2018. Evolution of dopamine receptors: phylogenetic evidence suggests a later origin of the DRD2l and DRD4rs dopamine receptor gene lineages. *PeerJ* 6:e4593. 10.7717/peerj.4593

Ott T, Nieder A. 2018. Dopamine and Cognitive Control in Prefrontal Cortex. *Trends in Cognitive Sciences*. 10.1016/j.tics.2018.12.006

Panneton WM. 2013. The mammalian diving response: an enigmatic reflex to preserve life? *Physiology (Bethesda, Md)* 28:284-297. 10.1152/physiol.00020.2013

Rambaut A, Drummond AJ. 2012. FigTree version 1.4.0.

Ramirez J-M, Folkow LP, Blix AS. 2007. Hypoxia Tolerance in Mammals and Birds: From the Wilderness to the Clinic. *Annual Review of Physiology* 69:113-143. 10.1146/annurev.physiol.69.031905.163111

Sadier A, Davies KTJ, Yohe LR, Yun K, Donat P, Hedrick BP, Dumont ER, Dávalos LM, Rossiter SJ, Sears KE. 2018. Multifactorial processes underlie parallel opsin loss in neotropical bats. *eLife* 7:e37412. 10.7554/eLife.37412

Sanada H, Xu J, Jose PA, Watanabe H, Felder RA. 2000. Differential expression and regulation of dopamine-1(d1) and dopamine-5(d-5) receptor function in human kidney. *American Journal of Hypertension* 13:156A-156A. 10.1016/S0895-7061(00)00649-X

Sharma V, Hecker N, Roscito JG, Foerster L, Langer BE, Hiller M. 2018. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications* 9:1215. 10.1038/s41467-018-03667-1

Somorjai IML, Martí-Solans J, Diaz-Gracia M, Nishida H, Imai KS, Escrivà H, Cañestro C, Albalat R. 2018. Wnt evolution and function shuffling in liberal and conservative chordate genomes. *Genome Biology* 19:98. 10.1186/s13059-018-1468-3

Springer MS, Murphy WJ. 2007. Mammalian evolution and biomedicine: new views from phylogeny. *Biological Reviews* 82:375-392. 10.1111/j.1469-185X.2007.00016.x

Strasser B, Mlitz V, Fischer H, Tschachler E, Eckhart L. 2015. Comparative genomics reveals conservation of filaggrin and loss of caspase-14 in dolphins. *Exp Dermatol* 24:365-369. 10.1111/exd.12681

Sunahara RK, Guan H-C, O'Dowd BF, Seeman P, Laurier LG, Ng G, George SR, Torchia J, Van Tol HHM, Niznik HB. 1991. Cloning of the gene for a human dopamine D5 receptor with higher affinity for dopamine than D1. *Nature* 350:614-619. 10.1038/350614a0

Tian R, Wang Z, Niu X, Zhou K, Xu S, Yang G. 2016. Evolutionary Genetics of Hypoxia Tolerance in Cetaceans during Diving. *Genome biology and evolution* 8:827-839. 10.1093/gbe/evw037

Tiberi M, Caron MG. 1994. High agonist-independent activity is a distinguishing feature of the dopamine D1B receptor subtype. *Journal of Biological Chemistry* 269:27925-27931.

Triposkiadis F, Karayannis G, Giamouzis G, Skoularigis J, Louridas G, Butler J. 2009. The Sympathetic Nervous System in Heart Failure: Physiology, Pathophysiology, and Clinical Implications. *Journal of the American College of Cardiology* 54:1747-1762. https://doi.org/10.1016/j.jacc.2009.05.015

Vanyukov MM, Moss HB, Gioio AE, Hughes HB, Kaplan BB, Tarter RE. 1998. An Association Between a Microsatellite Polymorphism at the DRD5 Gene and the Liability to Substance Abuse: Pilot Study. *Behavior Genetics* 28:75-82. 10.1023/A:1021463722326

Wang X, Villar VAM, Armando I, Eisner GM, Felder RA, Jose PA. 2008. Dopamine, kidney, and hypertension: studies in dopamine receptor knockout mice. *Pediatric nephrology (Berlin, Germany)* 23:2131-2146. 10.1007/s00467-008-0901-3

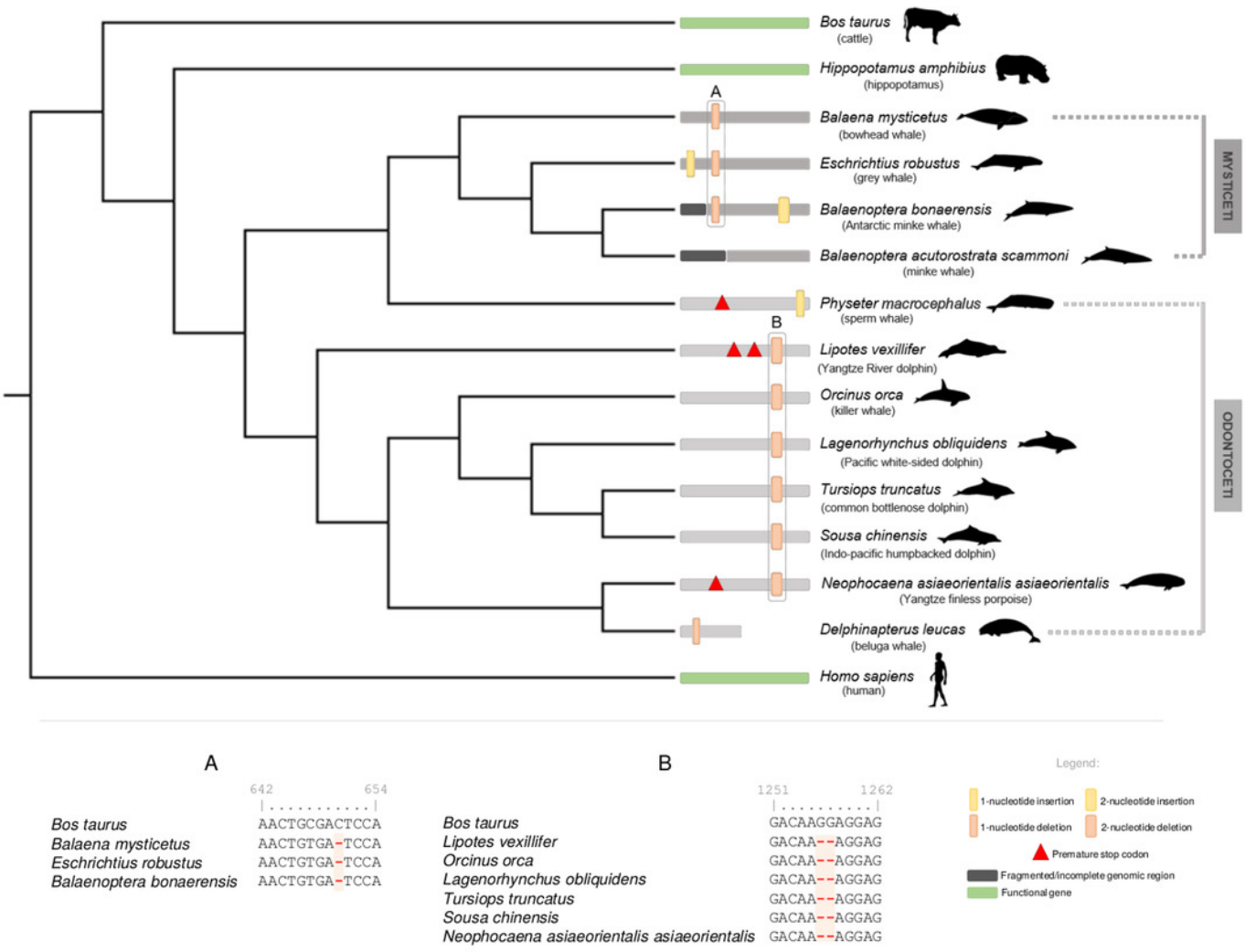
Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* 24:1586-1591. 10.1093/molbev/msm088

Zeng C, Yang Z, Wang Z, Jones J, Wang X, Altea J, Mangrum Amy J, Hopfer U, Sibley David R, Eisner Gilbert M, Felder Robin A, Jose Pedro A. 2005. Interaction of Angiotensin II Type 1 and D5 Dopamine Receptors in Renal Proximal Tubule Cells. *Hypertension* 45:804-810. 10.1161/01.HYP.0000155212.33212.99

Figure 1

Schematic representation of the DRD5 gene ORF abolishing annotated mutations regarding Cetacea parvorders Odontoceti and Mysticeti.

Phylogenetic relationships derived from a Maximum Likelihood (ML) tree with the DRD5 sequences from Mysticeti and Odontoceti species, as well as from *Hippopotamus amphibius*(hippopotamus), *Bos taurus*(cattle) and *Homo sapiens*(human). DRD_5 ORF-abolishing mutations are mapped in the corresponding branch. Vertical thick bars represent 2-nucleotide insertions (yellow) or deletions (orange). Vertical thin bars represent single nucleotide insertion (yellow) or deletion (orange). In-frame premature stop codons are represented by red triangles. Fragmented or incomplete genomic regions are represented by black regions. *Delphinapterus leucas*(beluga whale) predicted DRD_5 shortness is represented by a smaller bar. Functional branches are represented by green bars, nonfunctional branches are represented by grey and dark grey bars (Odontoceti and Mysticeti branches, respectively). Example of DRD_5 open reading frame (ORF) inactivating mutations concerning Mysticeti clade (A) and Odontoceti clade (excluding *Physeter catodon*, the sperm whale) (B). Numbers above characters represent the alignment position index.


Bos taurus
(cattle)
Hippopotamus amphibius
(hippopotamus)
A
Balaena mysticetus
(bowhead whale)
Eschrichtius robustus
(grey whale)
Balaenoptera bonaerensis
(Antarctic minke whale)
Balaenoptera acutorostrata scammoni
(minke whale)
MYSTICETI
Physeter macrocephalus
(sperm whale)
B
Lipotes vexillifer
(Yangtze River dolphin)
Orcinus orca
(killer whale)
Lagenorhynchus obliquidens
(Pacific white-sided dolphin)
Tursiops truncatus
(common bottlenose dolphin)
Sousa chinensis
(Indo-pacific humpbacked dolphin)
Neophocaena asiaeorientalis asiaeorientalis
(Yangtze finless porpoise)
Delphinapterus leucas
(beluga whale)
ODONTOCETI
Homo sapiens
(human)

A
642
654
Bos taurus AACTGCGACTCCA
Balaena mysticetus AACTGTGA-TCCA
Eschrichtius robustus AACTGTGA-TCCA
Balaenoptera bonaerensis AACTGTGA-TCCA
B
1251
1262
Bos taurus GACAAGGAGGAG
Lipotes vexillifer GACAA--AGGAG
Orcinus orca GACAA--AGGAG
Lagenorhynchus obliquidens GACAA--AGGAG
Tursiops truncatus GACAA--AGGAG
Sousa chinensis GACAA--AGGAG
Neophocaena asiaeorientalis asiaeorientalis GACAA--AGGAG
Legend:
1-nucleotide insertion
2-nucleotide insertion
1-nucleotide deletion
2-nucleotide deletion
Premature stop codon
Fragmented/incomplete genomic region
Functional gene

Table 1(on next page)

Selection (dN/dS) analyses with CODEML from PAML for the seven different branch categories.

Likelihood ratio tests (LRT) corresponding p -value concerning each branch category is also presented.

Table 1: Results of selection (d_N/d_S) analyses with CODEML from PAML for the seven different branch categories. Likelihood ratio test (LRT) corresponding p-value concerning each branch category is also presented.

Branch category	d_N/d_S **(ω)**
Functional branches (*Homo sapiens, Bos taurus, Hipoppotamus amphibius*)	0.09 (p = **0.030**)
Mysticeti branches (*Balaena mysticetus, Eschrichtius robustus, Balaenoptera bonaerensis, Balaenoptera acutorostrata scammoni*)	0.460 (p = 0.291)
Odontoceti branches (*Physeter macrocephalus*, *Lipotes vexillifer, Orcinus orca, Lagenorhynchus obliquidens, Tursiops truncatus, Sousa chinensis, Neophocanea asiaeorientalis asiaeorientalis*)	0.386 (p = 0.692)
Common cetacean branch	0.135 (p = **0.001**)
Stem Mysticeti branch	1.174 (p = 0.905)
Stem Odontoceti (excluding *Physeter macrocephalus*, sperm whale) branch	0.213 (p = **0.045**)
***Physeter macrocephalus* (sperm whale) ancestor branch**	0.320 (p = 0.460)

d_N/d_S values are significantly different from 1 if the correspondent LRT p-value < 0.05 (for a 95% confidence level, in bold).