Molecular evolution in dolphin hearing genes: different evolutionary pathways for marine and freshwater echolocation?

Magpali, L.; Freitas, L.; Ramos, E. K. S.; de Souza, E. M. S.; Nery, M. F.

Laboratório de Genômica Evolutiva. Departamento de Genética, Evolução, Microbiologia e Imunologia, Universidade Estadual de Campinas, Cidade Universitária, 13083970 - Campinas, SP - Brasil

Introduction

Echolocation was a key development in odontocete (i.e. toothed cetaceans) evolution, favouring their adaptive radiation through various habitats (Berta *et al.* 2015). In underwater environments, where vision is often limited, this sensory system allows dolphins to navigate, forage and understand the shapes and locations of objects through the emission, reception and processing of ultra-frequency sounds (Au 1993). Thus, the early emergence of specialized high-frequency hearing in odontocete ancestors was fundamental to the sonars of modern odontocetes (Ketten 1992; Park *et al.* 2016; Liu *et al.* 2018). High frequency hearing is mediated by the activity of inner and outer hair cells located in the auditory cochlea. Hair cell function and development are regulated by several known genes, involved in various pathways such as ion flow regulation (TMC1, CLDN14), motor activity for hearing adjustment (SLC26A5), structure integrity (CDH23), development and stereocilia movement (TMC1, SMPX) (Wang *et al.* 2020). Increasing evidence suggests that hearing genes are involved in functionally relevant pathways in echolocation and were targeted by natural selection in several dolphin lineages that evolved in different environments (Li *et al.* 2010; Liu *et al*. 2010; Davies *et al.* 2011; Shen *et al*. 2012; Liu *et al.* 2014; Costeur *et al.* 2018). The diversification of odontocetes allowed the evolution of several behavioral and physiological features, often generating convergent molecular patterns (Foote *et al.* 2015; Chikina *et al.* 2016). Echolocation, for instance, is employed by all dolphin species across different environments, regarding water depth, salinity and pressure, which constrain the propagation of sound underwater and thus might have influenced evolution of dolphin sonars. Recent studies argue that many phenotypical differences in echolocation among odontocetes were shaped by the different selective pressures of searching for food, navigating and communicating in oceanic, coastal and freshwater environments. In this sense, patterns of habitat variation in odontocetes have been correlated with morphological changes in the inner ear (e.g. cochlea shape) (Gutstein *et al.* 2014; Costeur *et al.* 2018; Park *et al.* 2019) and specializations in the acoustic parameters of the sonar (e.g. peak frequency, bandwidth) (Ketten 1992; Jensen *et al.* 2013; Ladegaard *et al.* 2015). In this context, we hypothesize that the distinct evolutionary pressures underlying the evolution of echolocation in marine and environments resulted in different patterns of molecular evolution, regarding positive selection and adaptive convergence, within hearing genes.

Materials and methods

Dataset. We selected five candidate hearing genes: CDH23, SLC26A5, TMC1, CLDN14, SMPX, and recovered their coding sequences from Genbank for 20 species of cetaceans (17 odontocetes and three baleen whales), along with approximately 70 external lineages per gene. Additionally, we assembled the coding sequences of *Sotalia guianensis* and *Sotalia fluviatilis* from newly sequenced whole genomes. To perform all following tests, we assembled the sequences into multi-species nucleotide alignments and translated them into amino acid alignments using PAL2NAL (Suyama et al. 2006), generating a total of 10 multi-species alignments organized in different datasets for each of the selection analyses.

Selection analyses. In order to identify individual branches and sites under positive selection, we performed branch model, branch-site model and site model analyses using the following methods nested on DataMonkey (Weaver *et al.* 2018): BUSTED (branch) and aBSREL (branch-site), to detect events of episodic diversifying selection throughout the genes (Murrell *et al.* 2015; Smith *et al.* 2015); as well as FUBAR and MEME, to detect individual sites under positive selection across all or some of the tested branches, respectively (Murrell *et al.* 2012, 2013). To test for differences in selective pressures at individual sites among sets of branches, we performed Contrast-FEL analysis with two sets of test branches, corresponding, respectively, to riverine and marine dolphins. Additionally, we carried branch model analysis as implemented on codeml, comparing the likelihood of three different models that estimate the ω ratio (equal to dN/dS, that is, the rate of nonsynonymous substitutions - dN - divided by the rate of synonymous substitutions - dS) for each branch in a phylogeny: one-model (single ω for all branches), two-model (two or more ω for each set of test branches, along with a background ω) and free-model (one ω for each branch).

Results and discussion

Branch-models. BUSTED detected overall signals of diversifying selection affecting some sites among all dolphins and among marine dolphins for TMC1 and CDH23; however, no similar signals were reported within the riverine dolphins group. Codeml analyses yielded higher support to the free-model for the genes TMC1, SLC25A5, CLDN14 and CDH23 and to the two-model for CDH23. In both tests, the likelihood values of the alternative hypotheses (free-model or two-model) were compared with the likelihood of the null hypothesis (one model) using a chi-square test to identify the model with the best significant likelihood support.

Table 2. Codeml results for the genes with evidence of different selection patterns among individual lineages. lnL = log likelihood of the corresponding model. LRT = 2 *[(lnL of the alternative model - lnL of the one model)], the alternative model being: (a) free model for the first LRT value of each gene, (b) two model for the second LRT value. np = number of parameters. df = degrees of frequency. p value corresponds to the statistical support obtained by chi-square assessments.

Gene	Model	omega (w)	lnL	LRT	np	df	p value
CDH23	one-model	0.05360	-120193.501770		128		
	free-model	various	-119593.429534	1200.144472	253	250	0
	two-model	0.05335; 0.10458	-120188.331	10.341578	129	1	0.0013
TMC1	one-model	8465	-22029.533014		98		
	free-model	various	-21762.825725	533.414578	193	95	0
	two-model	0.08449; 0.14409	-2202.28347	0.499082	99	1	1
SLC26A5	one-model	0.10684	-36087.322031		194		
	free-model	various	-35726.607205	721.429652	385	191	0
	two-model	0.10671; 0.27524	-36.086.695007	1.254048	195	1	1
CLDN14	one-model	0.07236	-12085.726900		152		
	free-model	various	-11900.725647	370.002506	301	149	0
	two-model	0.07242; 0.04992	-12.085.676615	0.10057	153	1	1

Branch-site models. aBSREL reported two marine lineages under episodic diversifying selection for CDH23: *Sotalia guianensis* (guiana dolphin), a coastal estuarine species from which the sister lineage *Sotalia fluviatilis* has recently diverged (Caballero *et al.* 2007) and *Physeter catodon* (sperm whale), a cetacean that is found exclusively in the open ocean and has developed a characteristic sonar possibly adapted to deep diving behavior (Tønnesen *et al.* 2020). SLC26A5 also presented one marine lineage under diversifying selection: *Phocoena sinus* (vaquita), a species of worldwide concern as the most endangered of extant cetaceans (Figure 1; Table 2). No evidence was found for episodic diversifying selection among the three exclusively riverine lineages, suggesting that the positive selection events among these lineages were weaker or fewer compared to marine dolphins, which was corroborated by the subsequent analyses below.

Table 2. Contrast-FEL results for differential selection among river and marine dolphins. The column R x M depicts the sites that showed significantly different ω ratios among these two groups. ω riverine, ω marine and ω background represent the ω values for these sites for each dolphin group and for the background lineages, respectively.

Branches under positive selection					
Name	**B**	**LRT**	**Test p-value**	**ω distribution over sites**	**#sites ω > 1**
Physeter catodon (CDH23)	0.0117	12.86	0.0044	ω1 = 0.00 (0.23%) ω2 = 0.196 (100%) ω3 = 100000 (0.21%)	7
Sotalia guianensis (CDH23)	0.0004	11.15	0.0091	ω1 = 0.00 (100%) ω2 = 1430 (0.11%)	4
Phocoena sinus (SLC26A5)	0.0000	9.98	0.0377	ω1 = 0.0236 (100%) ω2 = 475 (0.30%)	2

Site models. Both FUBAR and MEME (Table 4) recovered positive selected sites among dolphins, within four out of the five tested genes (exception: SMPX). Furthermore, we found different amounts of positively selected sites among riverine and marine dolphins, when analyzed separately with FUBAR (Table 4, highlighted numbers). Conversely, Contrast-FEL recovered significantly different ω ratios among these two groups, suggesting that some sites within CDH23, TMC1 and SLC26A5 have evolved under different selective pressures and natural selection rates between odontocetes from marine and freshwater environments (Table 3).

Table 3. Contrast-FEL results for differential selection among river and marine dolphins. The column R x M depicts the sites that showed significantly different ω ratios among these two groups. ω riverine, ω marine and ω background represent the ω values for these sites for each dolphin group and for the background lineages, respectively.

Gene	Total sites	Sites under positive selection				p-value
		R x M	*ω riverine*	*ω marine*	*dn/ds background*	
CDH23	3365	5	0.120	0.129	0.066	0.1
TMC1	766	1	0.133	0.132	0.110	0.1
SLC26A5	747	3	0.271	0.218	0.104	0.1
CLDN14	243	0	0.033	0.083	0.059	0.1
SMPX	86	0	0.453	0.457	0.142	0.1

Table 4. Positively selected sites detected by MEME and FUBAR. Ts = total sites screened for positive selection. pp = posterior probability. L and D are the numbers of positively selected sites for all lineages and dolphins, respectively.

Gene	ts	ps sites		p-value	All dolphins		River dolphins		Marine dolphins		
		L	*D*		*ω>1*	*ω<1*	*ω>1*	*ω<1*	*ω>1*	*ω<1*	*pp*
CDH23	3436	165	8	0.1	5	233	1	150	2	170	0.9
TMC1	812	51	1	0.1	2	24	2	20	1	29	0.9
SLC26A5	748	36	3	0.1	3	19	0	22	3	16	0.9
CLDN14	236	5	0	0.1	2	18	0	8	1	16	0.9
SMPX	86	2	0	0.1	0	1	0	0	0	1	0.9

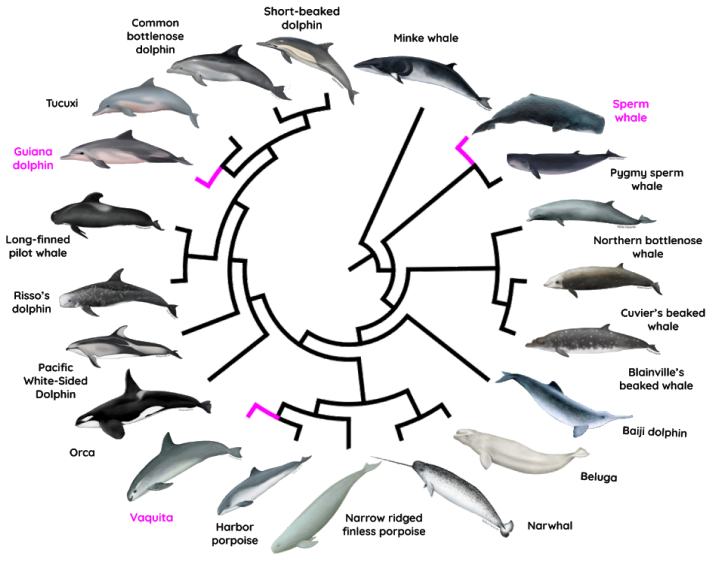

Figure 1. Phylogenetic tree, according to McGowen *et al.* (2020), of the extant dolphin lineages used in this study. Highlighted branches in magenta correspond to lineages under episodic diversifying selection recovered with aBSREL.

Taken together, our results reinforce the functional importance of hearing genes to echolocation, suggesting that river and marine dolphins might have undergone different evolutionary pathways regarding the molecular evolution of hearing genes, reflecting the contrasts between the two environments. Overall, positive selection is stronger and more abundant within marine dolphins, which might be correlated with higher selective constraints in these environments, since they are more variable than rivers in several physico-chemical conditions, such as temperature, depth and pressure.

References

Au, W. W. (1993). The sonar of dolphins. New York, Springer-Verlag.

Berta *et al.* (2015). Marine Mammals: Evolutionary Biology. Academic Press.

Caballero, S. *et al.* (2007). Taxonomic status of the genus Sotalia: species level ranking for "tucuxi"(Sotalia fluviatilis) and "costero"(Sotalia guianensis) dolphins. Marine mammal science, 23(2), 358-386.

Chikina, M. *et al.* (2016). Hundreds of genes experienced convergent shifts in selective pressure in marine mammals. Molecular biology and evolution, 33(9), 2182-2192.

Costeur, L. *et al.* (2018). The bony labyrinth of toothed whales reflects both phylogeny and habitat preferences. Scientific Reports, 8(1), 7841.

Davies, K. T. J. *et al.* (2012). Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. Heredity, 108(5), 480–489.

Foote, A. D. *et al.* (2015). Convergent evolution of the genomes of marine mammals. Nature genetics, 47(3), 272-275.z

Gutstein, C. S. *et al.* (2014). High frequency echolocation, ear morphology, and the marine–freshwater transition: A comparative study of extant and extinct toothed whales. Palaeogeography, Palaeoclimatology, Palaeoecology, 400, 62–74.

Jensen, F. H. *et al.* (2013). Clicking in shallow rivers: short-range echolocation of Irrawaddy and Ganges river dolphins in a shallow, acoustically complex habitat. PloS one, 8(4), e59284.

Ladegaard, M. *et al.* (2015). Amazon river dolphins (Inia geoffrensis) use a high-frequency short-range biosonar. The Journal of Experimental Biology, 218(Pt 19), 3091–3101.

Li, Y., Liu, Z. *et al.* (2010). The hearing gene Prestin unites echolocating bats and whales. Current Biology, 20(2), R55–R56.

Liu, Y. *et al.* (2010). Cetaceans on a molecular fast track to ultrasonic hearing. Current Biology, 20(20), 1834–1839.

Liu, Z. *et al.* (2014). Parallel sites implicate functional convergence of the hearing gene prestin among echolocating mammals. Molecular Biology and Evolution, 31(9), 2415–2424.

Liu, Z. *et al.* (2018). Genomic and functional evidence reveals molecular insights into the origin of echolocation in whales. Science advances, 4(10), eaat8821.

McGowen, M. R. *et al.* (2020). Phylogenomic resolution of the cetacean tree of life using target sequence capture. Systematic Biology, 69(3), 479-501.

Murrell, B. *et al.* (2012). Detecting individual sites subject to episodic diversifying selection. PLoS Genet, 8(7), e1002764.

Murrell, B. *et al.* (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Molecular biology and evolution, 30(5), 1196-1205.

Murrell, B. *et al.* (2015). Gene-wide identification of episodic selection. Molecular biology and evolution, 32(5), 1365-1371.

Park, T. *et al.* (2019). Convergent evolution in toothed whale cochleae. BMC evolutionary biology, 19(1), 195.

Parker, J.*et al.* (2013). Genome-wide signatures of convergent evolution in echolocating mammals. Nature, 502(7470), 228–231.

Shen, Y. Y. *et al.* (2012). Parallel evolution of auditory genes for echolocation in bats and toothed whales. PLoS Genetics, 8(6), e1002788.

Smith, M. D. *et al.* (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Molecular biology and evolution, 32(5), 1342-1353.

Suyama, M. *et al.* (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic acids research, 34(suppl_2), W609-W612.

Tønnesen, P. *et al.* (2020). The long-range echo scene of the sperm whale biosonar. Biology Letters, 16(8), 20200134.

Wang, H. *et al.* (2020). Evolutionary basis of high-frequency hearing in the cochleae of echolocators revealed by comparative genomics. Genome biology and evolution, 12(1), 3740-3753.

Weaver, S. *et al.* (2018). Datamonkey 2.0: a modernweb application for characterizing selective and other evolutionary processes. Molecular biology and evolution, 35(3), 773-777.

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution, 24, 1586–1591.