CrossMark

Selection on different genes with equivalent functions: the convergence story told by *Hox* genes along the evolution of aquatic mammalian lineages

Mariana F. Nery[1*], Brunno Borges[2], Aline C. Dragalzew[2] and Tiana Kohlsdorf[2*]

Abstract

Background: Convergent evolution has been a challenging topic for decades, being cetaceans, pinnipeds and sirenians textbook examples of three independent origins of equivalent phenotypes. These mammalian lineages acquired similar anatomical features correlated to an aquatic life, and remarkably differ from their terrestrial counterparts. Whether their molecular evolutionary history also involved similar genetic mechanisms underlying such morphological convergence nevertheless remained unknown. To test for the existence of convergent molecular signatures, we studied the molecular evolution of *Hox* genes in these three aquatic mammalian lineages, comparing their patterns to terrestrial mammals. *Hox* genes are transcription factors that play a pivotal role in specifying embryonic regional identity of nearly any bilateral animal, and are recognized major agents for diversification of body plans.

Results: We detected few signatures of positive selection on *Hox* genes across the three aquatic mammalian lineages and verified that purifying selection prevails in these sequences, as expected for pleiotropic genes. Genes found as being positively selected differ across the aquatic mammalian lineages, but we identified a substantial overlap of their developmental functions. Such pattern likely resides on the duplication history of *Hox* genes, which probably provided different possible evolutionary routes for achieving the same phenotypic solution.

Conclusions: Our results indicate that convergence occurred at a functional level of *Hox* genes along three independent origins of aquatic mammals. This conclusion reinforces the idea that different changes in developmental genes may lead to similar phenotypes, probably due to the redundancy provided by the participation of *Hox* paralogous genes in several developmental functions.

Keywords: Cetaceans, Pinnipeds, Sirenians, Molecular evolution, *Hox* genes, Positive selection

Background

The aquatic mammals – cetaceans (whales, dolphins and porpoises), pinnipeds (sea lions, seals and walruses) and sirenians (manatees and dugongs) – represent three extant mammalian lineages that independently recolonized the aquatic environment. Remarkably, these three groups encompass convergent anatomical and physiological solutions that accommodate the several challenges of

aquatic living. Their streamlined body shape minimizes drag, increases performance and reduces transport energetic costs [1]. Drag is also decreased by the presence of either a strikingly reduced pelvic appendicular skeleton or hindlimbs extremely modified and enlarged propulsive appendices [2]; these animals also evolved a pair of paddle-shapes fore-flippers, equivalent to the forelimbs of terrestrial mammals. In all cases, water recolonization involved extensive morphogenetic reorganization and modification of several physiological features [3]. Comparative anatomy and the fossil record have substantially elucidated these evolutionary transitions towards the

* Correspondence: mariananery@gmail.com; tiana@usp.br
[1]Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil
[2]Departamento de Biologia, FFCLRP, Universidade de São Paulo, Ribeirão Preto, Brazil

Nery *et al. BMC Evolutionary Biology* (2016) 16:113

Page 2 of 9

aquatic life (especially in cetaceans, see [4, 5]). But only recently, the molecular mechanisms underlying such dramatic and convergent morphological changes during evolution of these three mammalian lineages began to be unraveled (e.g. [6–13]).

The convergent evolution of similar traits in response to equivalent selective pressures represents an exceptional opportunity to study which evolutionary processes underlie specific phenotypic modifications associated to the conquest of a new environment. Several recent studies have been dedicated to evaluate the extent to which adaptive phenotypic convergence is attributable to convergent changes at the molecular level [14–17]. Such articles so far demonstrated that convergent evolution of similar phenotypes may or may not share similar molecular bases. Moreover, these similarities may occur at several genetic levels, meaning that the occurrence of convergence may involve the same mutations in the same genes, but may also comprise different and equivalent mutations in the same genes, or even mutations at different genes with equivalent molecular functions [14–17].

Molecular bases engaged in the phenotypic reorganization that occurred during the three evolutionary events of return to the aquatic environment by mammalian lineages likely comprise transcription factors essential for developmental processes. Among these, *Hox* genes are ideal candidates to test for associations between molecular evolution and morphogenetic rearrangements. These transcription factors regulate patterning of specific body structures during development by specifying regional identities along the anterior-posterior axis of all billaterian metazoans [18]. Vertebrates, in particular, possess four distinct *Hox* gene clusters (*Hox* A, B, C and D) located on different chromosomes, as a consequence of the two rounds of whole genome duplications that occurred early in their evolutionary history [19]. Their fundamental role in patterning body morphology, together with their broad contribution to developmental processes (e.g. [20–25]), foster the assertion that variation in nucleotide sequences of *Hox* gene coding regions is usually associated to phenotypic differences [26–31]. As a consequence, this gene family is often recognized as a major agent for diversification of metazoan body plans [19, 20]. Evaluation of molecular evolutionary patterns of *Hox* genes in the context of independent origins of equivalent phenotypes, however, remains relegated [but see [32] for a recent study on the topic].

In this study we test for convergent signatures in *Hox* genes by comparing sequences of three mammalian lineages that are phenotypically very similar: cetaceans, pinnipeds and sirenians. Our goal is to provide a more comprehensive understanding of *Hox* genes molecular

evolution in these aquatic lineages that secondarily and independently recolonized the sea, by characterizing and comparing the coding regions of *Hox* genes in aquatic and terrestrial mammals. Specifically, we test whether *Hox* genes underwent different selective pressures in aquatic mammals, in nucleotide and amino acid sequences, and also whether there are molecular signatures of these genes that endorse convergent developmental mechanisms during evolutionary transitions towards re-colonization of aquatic environments.

Methods

In order to accomplish a wide taxonomic sampling, we obtained unannotated genomic sequences in Ensembl genome browser and in GenBank from the following 19 mammalian species: tasmanian devil (*Sarcophilus harrisii*), microbat (*Myotis lucifugus*), elephant (*Loxodonta africana*), manatee (*Trichechus manatus*), horse (*Equus caballus*), dog (*Canis familiaris*), ferret (*Mustela putorius*), Weddell seal (*Leptonychotes weddellii*), walrus (*Odobenus rosmarus*), cow (*Bos taurus*), bottlenose dolphin (*Tursiops truncatus*), baiji dolphin (*Lipotes vexilifer*), minke whale (*Balaenoptera acutorostrata*), mouse (*Mus musculus*), rabbit (*Oryctologus cuniculus*), human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), orangutan (*Pongo abelli*) and marmoset (*Callithrix jacchus*). Despite the low genome coverage of some species, the four *Hox* clusters (A, B, C and D) have been well sequenced, and were complete and located within single scaffolds for most species, excepting the Weddell seal, which cluster A was incomplete, cluster B was fragmented in two scaffolds, and clusters C and D were absent, the microbat, which had the Cluster B fragmented in two scaffolds, and the baiji dolphin, which Cluster C remained absent of databases.

Genomic sequences of *Hox* genes from these species were manually annotated based on comparisons with known exon sequences available for humans, using the programs GenScan [33] and Blas2seq 2.2 [34]. To further corroborate the orthologous relationships of each gene included in our analyses, we conducted phylogenetic reconstructions based on maximum likelihood, using the program RAxML [35]. Moreover, in order to maximize the number of orthologs in each analysis, we conducted a search in GenBank for *Hox* genes from other mammalian species. Using this approach, the species composition of each orthologous *Hox* gene varied from 11 up to 38 species. It means that each *Hox* gene had a unique phylogenetic tree, comprising the ortholog specific set of species recovered from sequenced genomes or GenBank. After this verification of orthologous sequences we finally proceeded with the natural selection analyses. All accession numbers and species included in each *Hox* gene analysis are depicted in Additional file 1: Table S1. All nucleotide sequences

Nery *et al. BMC Evolutionary Biology* (2016) 16:113

Page 3 of 9

were aligned using the program PRANK [36], and nucleotide alignments were generated using the amino acid alignments as a template in the software PAL2NAL [37].
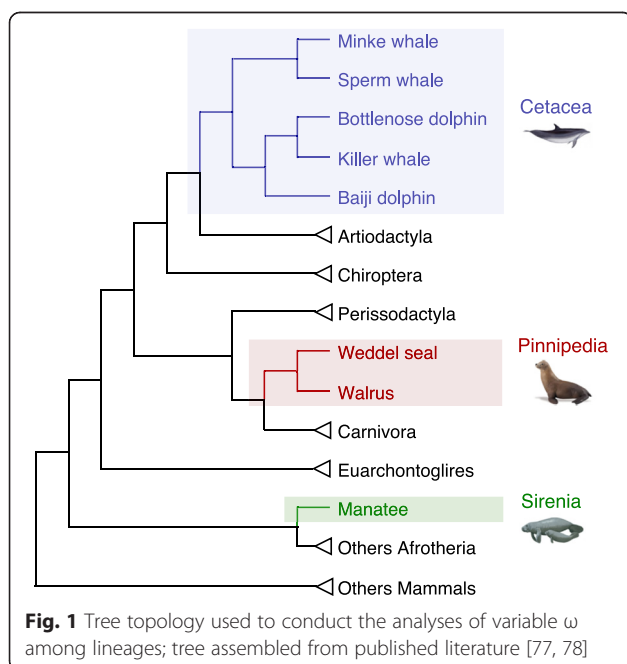
In order to investigate the possible role of changes in evolutionary rates and to test for positive selection on *Hox* genes evolution in aquatic mammalian lineages, we used the maximum likelihood codon substitution model implemented in the PAML 4.7 package [38]. The following branch-models of variable ω (= d_N/d_S) rates among lineages were implemented, based on Fig. 1: the *one-ratio model* that assigns the same ω ratio for all branches, and a *two-ratio model*, where one ω was assigned for the ancestral branch of Cetacea, Pinnipedia and Sirenia, treated separately, and another ω was assigned for all remaining mammals. This approach allowed us to identify whether any *Hox* gene in an aquatic lineage exhibits accelerated evolutionary rates when compared to terrestrial counterparts.

We also aimed to detect evidence of positive selection on specific sites along a specific lineage, so we applied branch-site models in PAML, which compare changes in ω along the different sites in the alignment between the foreground branch of primary interest and the background branches [39]. Branches leading to the ancestor of the cetaceans, to the ancestor of pinnipeds, and to the sirenian lineage were chosen as foreground branches in each analysis. We compared the modified model A [40], in which some sites are allowed to change to an ω > 1 in the foreground branch, with the corresponding null hypothesis of neutral evolution. The Bayes Empirical Bayes (BEB) method identified sites under positive selection [40]. In all cases, three starting ω values (0.5, 1.0 and

2.0) were used to verify the existence of multiple local optima. Nested models were compared using the likelihood ratio test (LRT), and the level of significance was settled at 0.05.

Branch-site models may be limiting because they require the prior identification of foreground lineages and the assumption that $d_N/d_S = 1$ for all background lineages [41]. For that reason, we examined each *Hox* gene for signatures of episodic positive selection using a *mixed effects model of evolution* [MEME, 42], performed with HyPhy package implemented in DataMonkey Web Server [43]. The best-fitting nucleotide substitution model was selected through the automatic model tool available on the server, and then ω variation was evaluated among sites and lineages simultaneously. In doing so, this model allows identification of episodes of positive selection that affect only a subset of lineages, which are often ignored using other methods assuming that the ω value is shared by all sites in the alignment or that selective pressures are constant throughout time [42]. In MEME there is no need to specify foreground or background branches. As a complementary approach, we also used another method named BUSTED (Branch-site Unrestricted Statistical Test for Episodic Evolution [44]), implemented in the DataMonkey web server. This method is capable of detecting positive selection that has acted on a subset of branches in a phylogeny at a subset of sites within a gene, and the foreground branch has to be indicated. Sites positively selected were identified at a significance level of P <0.05.

As a final step, we also inferred possible physicochemical differences in *Hox* sequences among the mammalian lineages studied. The aforementioned codon-based analyses implemented in PAML and in DataMonkey do not take into account the magnitude of changes in the physicochemical properties of amino acids resulting from a nonsynonymous substitution. To detect significant physicochemical amino acid changes among residues in *Hox* genes, we used therefore the algorithm implemented in TreeSAAP 3.2 software [45]. This program compares the magnitude of property changes of non-synonymous residues inferred from a phylogeny, and indicates which amino acid properties likely have been affected by positive destabilizing selection during the evolutionary process. In TreeSAAP, the magnitudes of nonsynonymous changes are classified into eight categories according to the change in specific physicochemical properties, from conservative (1–3) to very radical substitutions (6–8). For each category, a *z-score* is calculated. Significant positive *z-scores* indicate that a given region is under influence of positive selection (i.e., the number of inferred amino acid replacements significantly exceeds the number of those expected by chance). Here we only considered significant substitutions

Fig. 1 Tree topology used to conduct the analyses of variable ω among lineages; tree assembled from published literature [77, 78]

Nery *et al. BMC Evolutionary Biology* (2016) 16:113

Page 4 of 9

assigned to the highest categories representing extreme changes in physicochemical properties (categories 7 and 8) at the $P < 0.001$ level. The physicochemical properties identified in these categories were then subjected to a sliding window analysis of 20 codons in width to verify which specific regions in the protein differ significantly from a neutral model.
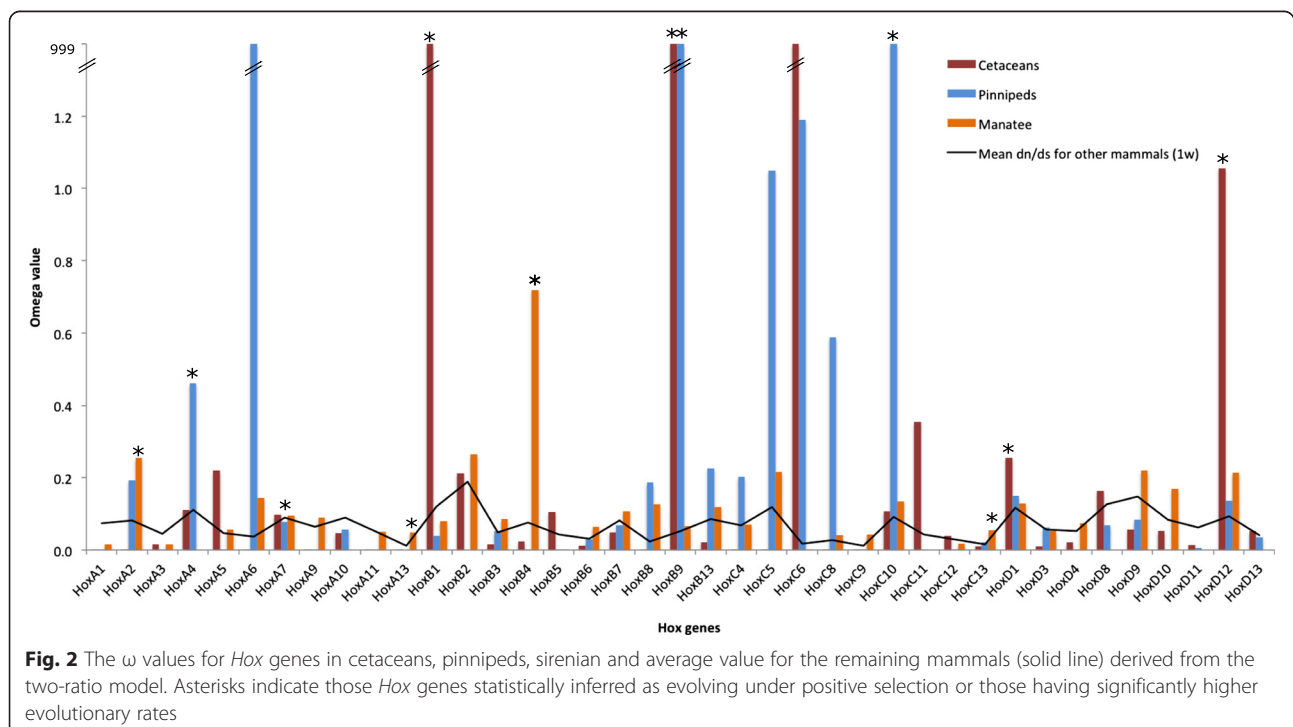
Results and discussion

Cetaceans, pinnipeds and sirenians have independently evolved very similar specialized body plans and associated physiological features for an aquatic lifestyle. Molecular mechanisms underlying these key adaptations have recently received increasing interest [9], and the question of whether such convergent phenotypic adaptations share similar molecular bases deserves special attention due to the amount of complete sequenced genomes newly available for several mammals [46, 47]. Similarity may occur at many levels, such as nucleotides, genes, networks and functions [14]. In agreement with the postulate that *Hox* molecular evolution extensively engages major morphological transitions in vertebrates, our results from different analyses focusing on selective regimes acting during *Hox* genes evolution suggest that independent evolution of phenotypically similar aquatic mammalian lineages involved developmental convergence residing on *Hox* functions instead of on *Hox* genes, as further detailed.

Our evaluation of *Hox* genes evolution in aquatic mammals started with annotation of the full complement of structural genes in the *Hox* clusters on unannotated genomic sequences from 19 mammalian species representatives from all major mammalian lineages, an approach where we also included known *Hox* sequences of additional mammalian species retrieved from GenBank. We surveyed all the 39 *Hox* protein sequences, which represent all the gene family members. As expected, synteny is conserved in all species and the cluster size was very similar among lineages (Additional file 2: Table S2). Maximum likelihood phylogenies arranged *Hox* genes into well-supported clades (data not shown), allowing us to establish orthologous relationships of the manually annotated genes.

We aimed to detect possible roles of positive selection during the evolution of *Hox* genes in aquatic mammalian lineages, so we implemented different models using a maximum likelihood approach in the program PAML and in the program HyPhy within DataMonkey server. As summarized in Fig. 2 (results from the branch models implemented in PAML) and the Additional file 3: Table S3 (likelihood values for all models), several *Hox* genes have evolved under specific selective regimes in one or more mammalian aquatic lineages. The one-ratio model showed that the ω values (ranging from 0.011 in *HoxC9* to 0.189 in *HoxB2*) were significantly lower than 1. These low values were already expected because *Hox* genes, which are known to be pleiotropic, have experienced constrained selective pressures to maintain their function. We then applied the two-ratio model for each *Hox* gene, in order to estimate one ω value

Fig. 2 The ω values for *Hox* genes in cetaceans, pinnipeds, sirenian and average value for the remaining mammals (solid line) derived from the two-ratio model. Asterisks indicate those *Hox* genes statistically inferred as evolving under positive selection or those having significantly higher evolutionary rates

Nery *et al. BMC Evolutionary Biology* (2016) 16:113

Page 5 of 9

separately for Cetacea, Sirenia and Pinnipedia and another ω to all the remaining terrestrial mammals. The LRT tests showed that the two-ratio model fits significantly better than the one-ratio model for *HoxB1*, *HoxB9*, *HoxD1* and *HoxD12* genes in the cetacean lineage, for *HoxA4*, *HoxB9* and *HoxC10* genes in the pinnipedian lineage, and for *HoxA2*, *HoxA13*, *HoxB4* and *HoxC13* genes in the sirenian lineage (Table 1), suggesting that these genes were subjected to different selective pressures during their evolution when compared to other mammalian counterparts. The higher ω values for these genes suggest significantly accelerated rates of evolution in these lineages. Such accelerated rates could either represent relaxation on purifying selection or reflect the action of positive selection. These results suggest that evolution of equivalent phenotypes may have recruited changes in *Hox* genes that were specific in each group.

Branch models estimate an overall ω value for all codons in the gene, but typically positive selection acts only on a few sites and within a short evolutionary time period [48]. Therefore, we have also explored the variation of evolutionary rates in specific amino acid sites using branch-site models, BUSTED and MEME methods. The branch-site model only identified sites under positive selection in Cetacea (*HoxB1*) and Sirenia (*HoxC13*), while the BUSTED and MEME methods identified episodic positive selection in Cetacea (*HoxB1*), Pinnipedia (*HoxA7*) and Sirenia (*HoxC13*), as shown in Table 1. Differences from the results of each model can be attributed to differences in their underlying assumptions [49]. However, the overlapping results among the three models provides good evidence that these genes were indeed subject to non-neutral selective pressures. For example, both *HoxB1* in cetaceans and *HoxC13* in sirenians were identified as being under positive selection by all models.

Overall, the signatures of positive selection are much less prevalent across *Hox* genes than those of purifying selection; they are concentrated in few genes and seem exclusive to each aquatic mammalian lineage. The gene *HoxB9* was the only gene found as positively selected in
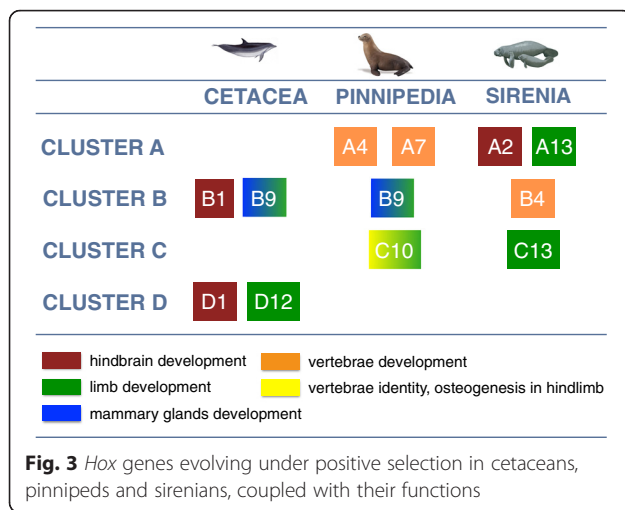
two (cetaceans and pinnipeds) out of the three lineages of aquatic mammals, and none *Hox* gene was inferred as being under positive selection in all three lineages. The apparent lack of a convergence (i.e., different genes being positively selected), however, is revisited when we evaluate the developmental roles of these genes (Fig. 3). The genes *HoxB1*, *HoxD1* (cetaceans) and *HoxA2* (sirenians) are all involved in hindbrain development [50–53]. These are part of the most anterior paralogous groups 1, 2 and 3 that are expressed early during ontogeny and in the anterior axis of the body [54]. Also, the genes *HoxA4*, *HoxA7* (pinnipeds) and *HoxB4* (sirenians) overlap roles related to vertebral development [55–57]. Finally, the genes *HoxB9* (cetaceans and pinnipeds), *HoxD12* (cetaceans), *HoxA13* and *HoxC13* (sirenians) are essential for limb and digit development [58–60]. The gene *HoxC10* was only identified as having accelerated evolutionary rates in pinnipeds; this gene regulates vertebral identity at the transition from thoracic to lumbar and lumbar to sacral regions, and also plays a general role regulating chondrogenesis and osteogenesis in the hindlimb [61]. It is noteworthy that this specific gene had an accelerated evolution rate only in pinnipeds, because this lineage underwent major modifications of their hindlimbs, instead of reduction or loss, which were considerably modified as an adaptation to an amphibian life, since they still use land to reproduce and breed [2].

The substantial overlap among developmental functions of the genes positively selected in aquatic mammals may be interpreted in the light of the history of this complex and unique gene family. It is widely accepted that an ancestral *Hox* gene cluster was duplicated twice during the emergence of vertebrates, as a result of two rounds of whole genome duplications [19, 23]. Consequently, paralogous *Hox* genes, those located at the same relative positions within each of the four clusters, reveal high sequence similarities and shared properties [62]. After duplications, there were secondary losses, and each cluster has selectively retained different subsets of paralogous genes. Those that remained show a strong functional complementarity [63]. The overlapping and

Table 1 Genes and sites inferred to be under positive selection using different models implemented in PAML

Model	Cetacea	Positively selected sites	Pinniped	Positively selected sites	Sirenia	Positively selected sites
Branch model	**HoxB1**, HoxB9, HoxD1, HoxD12	-	HoxA4, HoxB9, HoxC10	-	HoxA2, HoxA13, HoxB4, **HoxC13**	-
Branch-site model	**HoxB1**	18, 35, 64, 147, 153, 169			**HoxC13**	16, 28, 200, 233
MEME	**HoxB1**	64	HoxA7	201	**HoxC13**	233
BUSTED	**HoxB1**	18, 35, 64, 128, 147, 153, 169	HoxA7	201	**HoxC13**	233

In bold are highlighted those genes identified by all methods

Nery *et al. BMC Evolutionary Biology* (2016) 16:113

Page 6 of 9

Fig. 3 *Hox* genes evolving under positive selection in cetaceans, pinnipeds and sirenians, coupled with their functions

complementing functions of the paralogous genes may provide different possible evolutionary routes that achieve the same phenotypic solution for equivalent challenges. In other words, different possible combinations of molecular mechanisms may lead to convergence in trait function, not necessarily through the same molecular solution (i.e. same amino acid substitution in the same gene).

The few studies focusing on *Hox* genes in aquatic mammals relate to cetaceans, and none undergone a comparative analyses of the three aquatic lineages. From these studies, we observed for example that the gene *HoxD12*, previously identified as evolving under positive selection in dolphins [64], was also identified in our analysis as having a significant higher ω value in cetaceans, although we did not identify any particular site evolving under positive selection. This gene is strongly expressed in developing limb buds and is highly conserved across vertebrates [65], but the cetacean sequences are considerably variable when compared to other vertebrates (data not shown here). Also, recent sequence analyses of members from the *Hox* gene family in cetaceans [66] identified significant positive selection only in the gene *HoxB9* in dolphins, at site 175. Our results, just like with *HoxD12*, identified *HoxB9* as having signatures of accelerated evolutionary rates in cetaceans and pinnipeds, but no site was detected as positively selected. As already pointed out by [66], all four paralogous *Hox9* genes act in concert to establish the forelimb posterior domain by regulating *Hand2* expression in this region [67]. Hence, the modified *HoxB9* in cetaceans and pinnipeds might have contributed for acquisition of their fore-flipper. The conflicting results among our analyses and those from previous studies [66, 67] likely reside on the fact that *Hox* genes are subjected to strong long-term purifying selection, which in turn may have hidden any short-term positive selection signal [43]. Therefore, differences

in tree topologies and models used for analyses of selective pressures might influence the results. Differently from the other two previous works [66, 67], which used only the dolphin as a cetacean representative and partial coding sequences, here we included complete sequences for all species considered, and always had more than one cetacean representative in our analyses, besides using several models with different assumptions. Such methodological design endorsed accuracy and strictness to our study.

Recently, two independent studies used genomic approaches to investigate the molecular basis of the convergent evolution of cetaceans, pinnipeds and sirenians [46, 47]. The genomic analyses from [46] found that convergent amino acid substitutions were widespread throughout the genome and that a subset of these substitutions was in genes evolving under positive selection. They concluded that convergent evolution likely most arises from different molecular pathways to reach the same phenotypic outcome. Likewise, [47] found convergence in protein coding genes along the whole genome associated with aquatic lifestyle is mainly characterized by independent substitutions and relaxed purifying selection. Both genomic studies [46, 47] did not focus specifically on the *Hox* gene family or assessed the possible functional convergence among genes identified as being positively selected. But in accordance to our results, they propose that sequence convergence in aquatic mammals is predominantly characterized by independent rather than parallel substitutions.

In the past decades, research on molecular mechanisms of convergent phenotypes has expanded greatly, and the results have demonstrated that convergent phenotypic evolution may be attributable to similar molecular bases [14, 15, 17]. Such similarities may however occur at several genetic levels, meaning that phenotypic convergent adaptation might emerge through identical convergent mutations (e.g. [68–70]) but may also comprise numerous alternative pathways (e.g. [71–73]). Regarding such alternative pathways, recent studies even suggest that convergence would be more properly accessed through evaluation of gene functions and functional complexes as selection ultimately targets multigenic functional groups instead of unique amino acid sites [74, 75]. Similarly to our results, these studies demonstrate equivalent biological processes undergoing accelerated evolution in phenotypically similar animals, although through changes in lineage-specific sets of genes [71], adding evidence to the statement that convergent evolution involves a mosaic of molecular changes. Moreover, as pointed out before by [16], expanding our view to the several possible levels of the convergence spectrum has the potential to reveal new insights about the mechanisms underlying phenotypic convergence.

Nery *et al. BMC Evolutionary Biology* (2016) 16:113

Page 7 of 9

The relevance of *Hox* genes for phenotypic evolution endorses discussion about the functional implications of the amino acid properties that are under selection. Our TreeSAAP analyses, which compare the observed distribution of physicochemical changes inferred from a phylogenetic tree with an expected distribution based on the assumption of completely random amino acid replacement that is expected under the conditions of selective neutrality [45, 76], provide evidence for significant physicochemical amino acid changes among residues in *Hox* genes of aquatic mammalian lineages. Similar to our results of the codon models, negative selection dominates the categories of radical changes, but significant physicochemical amino acid changes were detected in all *Hox* genes identified previously as having sites under positive selection or showing significantly accelerated rates of evolution. Evidence for positive selection was recognized in five physicochemical properties, with global *z-scores* above the significance threshold (z > 3.09). These *Hox* genes exhibited at least one physicochemical property under positive destabilizing selection, which is known to interfere both at chemical and structural levels. The properties were: alpha-helical tendencies (*HoxA2*, *HoxA4*, *HoxA13*, *HoxB1*, *HoxB4*, *HoxC10*, *HoxC13* and *HoxD1*), power to be at the C-terminal (*HoxA4*, *HoxB1*), turn tendencies (*HoxA2*, *HoxB1*, *HoxB4*), coil tendencies (*HoxB1*, *HoxB4*) and thermodynamic transfer hydrophobicity (*HoxD12*) (Table 2, Additional file 4: Figure S1). The positive-destabilizing selection acting on such properties in these genes might contribute for adaptive evolution by influencing the gene biochemical and conformational character.

Identifying molecular signatures in the *Hox* gene family in three mammalian lineages that independently recolonized aquatic environments highlights the role of changes in developmental functions during recurrent evolution of similar phenotypes. It is important, however, to note that aquatic adaptations of mammals likely involve both changes on coding genes and on elements regulating expression patterns. Whether changes in *Hox* expression patterns contributed more to morphological adaptations in aquatic mammals than changes in *Hox*-protein-function remains to be answered. We can, though, assure that at least some of the *Hox* members belonging to this important gene family were subjected to positive selection during the evolution of cetaceans, pinnipeds and sirenians, and very likely played an important role in the evolution of aquatic adaptations. More importantly, our study kindles the convergence of *Hox* gene developmental functions as a major factor underlying independent evolution of similar phenotypes in three mammalian aquatic lineages. In this way, our results contribute to the ongoing discussion regarding the molecular basis of convergence, adding information to the growing body of evidence that indicate that convergence is hierarquical and may occur at several biological levels.

Conclusions

This study provides a detailed characterization of the pattern of selection pressures for the entire *Hox* gene family in mammals, and analyzes the extent of positive selection events on these genes with special focus on three aquatic lineages. All *Hox* genes investigated here experienced strong purifying selection, suggesting a conservative general evolutionary pattern in mammals, and positive selection affects only a specific fraction of sites. However, there is substantially more overlap at the level of their developmental functions than of their nucleotide sequences, reflecting the functional complexity of the *Hox* gene family, which is composed by paralogous groups having complimentary functions. Because of their evolution mode (i.e. four extant clusters originated from two events of duplication of only one ancestral cluster), the *Hox* gene family appears to be relatively 'loose' in the sense that distinct lineages exhibit convergent

Table 2 Amino acid physicochemical properties under positive destabilizing selection in *Hox* genes

Genes	Alpha-helical tendencies	Turn tendencies	Power to be at the N-terminal	Coil tendencies	Thermodynamic transfer hydrophobicity
HoxA2	x	x			
HoxA4	x		x		
HoxA7					
HoxA13	x				
HoxB1	x	x	x	x	
HoxB4	x	x		x	
HoxC10	x				
HoxC13	x				
HoxD1	x				
HoxD12					x

Nery et al. BMC Evolutionary Biology (2016) 16:113

Page 8 of 9

molecular evolution involving similar developmental functions that are not settled on the exact same genes.

Availability of data and material

The accession numbers from Ensembl and GenBank of the dataset supporting the conclusions of this article are included within Additional file 1: Table S1.

Additional files

Additional file 1: Table S1. Accession numbers of all mammalian species included in our study and the total number of species included in each *Hox* gene analysis. (XLSX 64 kb)

Additional file 2: Table S2. *Hox* clusters size of those mammals with sequenced genomes included in our study. (DOCX 114 kb)

Additional file 3: Table S3. Log likelihhod (lnL), number of parameters (np), omega values (ω) and Likelihood Ratio Test (LRT) values under different models for each *Hox* gene. (DOCX 146 kb)

Additional file 4: Figure S1. Sliding window plots of the *z-scores* of radically transitions of amino acid properties showing protein regions under positive destabilizing selection in *Hox* genes. (PDF 394 kb)

Abbreviations

BEB: Bayes Empirical Bayes; LRT: likelihood ratio test; MEME: mixed effects model of evolution; BUSTED: Branch-site Unrestricted Statistical Test for Episodic Evolution.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MFN conceived the study. MFN, BB and ACD analyzed the data. MFN and TK wrote the manuscript. All authors read and approved the final manuscript.

References

1. Berta A, Sumich JL, Kovacs KM. Marine Mammals Evolutionary Biology. 2nd ed. San Diego: Academic Press; 2006.
2. Hoelzel AR. Marine mammal biology. An evolutionary approach. Cornwall: Blackwell Publishing; 2002.
3. Gingerich PD. Rates of evolution: effects of time and temporal scaling. Science. 1983;222:159–61.
4. Uhen MD. Evolution of marine mammals: Back to the sea after 300 million years. Anat Rec. 2007;290(6):514–22.
5. Uhen MD. The origin(s) of whales. Annu Rev Earth Planet Sci. 2010;38:189–219.
6. Nery MF, González DJ, Opazo JC. How to make a dolphin: molecular signature of positive selection in cetacean genome. PLoS One. 2013;8(6):e65491.
7. Nery MF, Arroyo JI, Opazo JC. Accelerated evolutionary rate of the myoglobin gene in long-diving whales. J Mol Evol. 2013;76(6):380–7.
8. Nery MF, Arroyo JI, Opazo JC. Genomic organization and differential signature of positive selection in the alpha and beta globin gene clusters in two cetacean species. Genome Biol Evol. 2013;5(12):2359–67.
9. McGowen MR, Gatesy J, Wildman DE. Molecular evolution tracks macroevolutionary transitions in Cetacea. Trends Ecol Evol. 2014;29(6):336–46.
10. Yim HS, Cho YS, Guang X, Kang SG, Jeong JY, Cha SS, Oh HM, Lee JH, Yang EC, Kwon KK et al. Minke whale genome and aquatic adaptation in cetaceans. Nat Genet. 2014;46(1):88–92.
11. Yim H-S, Cho Y, Guang X, Kang S, Jeong J-Y, Cha S-S, et al. Minke whale genome and aquatic adaptation in cetaceans. Nat Genet. 2013;46:88–92.
12. Zhou X, Sun F, Xu S, Fan G, Zhu K, Liu X, Chen Y, Shi C, Yang Y, Huang Z et al. Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptation. Nat Commun. 2013;4:2708.
13. Wang J, Yu X, Hu B, Zheng J, Xiao W, Hao Y, Liu W, Wang D. Physicochemical evolution and molecular adaptation of the cetacean osmoregulation-related gene UT-A2 and implications for functional studies. Sci Rep. 2015;5:8795.
14. Rosenblum EB, Parent CE, Brandt EE. The Molecular Basis of Phenotypic Convergence. Annu Rev Ecol Evol Syst. 2014;45(1):203–26.
15. Stern D. The genetic causes of convergent evolution. Nat Rev Genet. 2013;14:751–64.
16. Manceau M, Domingues V, Linnen C, Rosenblum E, Hoekstra H. Convergence in pigmentation at multiple levels: mutations, genes and function. Philos Trans R Soc Lond B Biol Sci. 2010;365:2439–50.
17. Storz J. Causes of molecular convergence and parallelism in protein evolution. Nat Rev Genet. 2016;17:239–50.
18. McGinnis W, Krumlauf R. Homeobox genes and axial patterning. Cell. 1992;68:283–302.
19. Liang D, Wu R, Geng J, Wang C, Zhang P. A general scenario of Hox gene inventory variation among major sarcopterygian lineages. BMC Evol Biol. 2011;11(1):25.
20. Carroll SB. Chance and necessity: the evolution of morphological complexity and diversity. Nature. 2001;409:1102–9.
21. Bejder L, Hall BK. Limbs in whales and limblessness in other vertebrates: mechanisms of evolutionary and developmental transformation and loss. Evol Dev. 2002;4(6):445–58.
22. Hoegg S, Meyer A. Hox clusters as models for vertebrate genome evolution. Trends Genet. 2005;21(8):421–4.
23. Kuraku S, Meyer A. The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. Int J Dev Biol. 2009;53(5-6):765–73.
24. Mallo M, Wellik DM, Deschamps J. Hox genes and regional patterning of the vertebrate body plan. Dev Biol. 2010;344(1):7–15.
25. Woltering JM, Noordermeer D, Leleu M, Duboule D. Conservation and divergence of regulatory strategies at Hox loci and the origin of tetrapod digits. PLoS Biol. 2014;12(1):e1001773.
26. Goodman FR. Limb malformations and the human HOX genes. Am J Med Genet. 2002;112(3):256–65.
27. Ronshaugen M, McGinnis N, McGinnis W. Hox protein mutation and macroevolution of the insect body plan. Nature. 2002;415:914–7.
28. Innis JW, Goodman FR, Bacchelli C, Wagner GP. A HoxA13 allele with a missense mutation in the homeobox and a dinucleotide deletion in the promoter underlies Guttmacher syndrome. Hum Mutat. 2002;19:573–4.
29. Fares MA, Bezemer D, Moya A, Marín I. Selection on coding regions determined Hox7 genes evolution. Mol Biol Evol. 2003;20(12):2104–12.
30. Brayer KJ, Lynch VJ, Wagner GP. Evolution of a derived protein-protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. Proc Natl Acad Sci U S A. 2011;108:E414–20.
31. Hao Z, Dai J, Shi D, Xu Z, Chen D, Zhao B, Teng H, Jiang Q. Association of a single nucleotide polymorphism in HOXB9 with developmental dysplasia of the hip: a case-control study. J Orthop Res. 2013;32(2):179–82.
32. Singarete ME, Grizante MB, Milograna SR, Nery MF, Kin K, Wagner GP, Kohlsdorf T. Molecular evolution of HoxA13 and the multiple origins of limbless morphologies in amphibians and reptiles. Genet Mol Biol. 2015;38(3):255–62.
33. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 1997;268:78–94.
34. Tatusova TA, Madden TL. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett. 1999;174:247–50.
35. Stamatakis A. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
36. Loytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A. 2005; 102(30):10557–62.
37. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Bioinformatics. 2006;34:w609–12.
38. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.

Nery *et al. BMC Evolutionary Biology* (2016) 16:113

Page 9 of 9

39. Zhang J, Nielsen R, Yang Z. Evaluation of an Improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 2005;22(12):2472–9.

40. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22:1107–18.

41. Pond SL, Murrell B, Fourment M, Frost SD, Delport W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol. 2011;28(11):3033–43.

42. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8(7):e1002764.

43. Pond SLK, Frost SDW. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol. 2005;22(3):478–85.

44. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM et al. Gene-wide identification of episodic selection. Mol Biol Evol. 2015;32(5):1365–71.

45. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. TreeSAAP: Selection on Amino Acid Properties using phylogenetic trees. Bioinformatics. 2003;19(5):671–2.

46. Foote AD, Liu Y, Thomas GW, Vinar T, Alföldi J, Deng J, Dugan S, Elk CE, Hunter ME, Joshi V et al. Convergent evolution of the genomes of marine mammals. Nat Genet. 2015;47:272–5.

47. Zhou X, Seim I, Gladyshev VN. Convergent evolution of marine mammals is associated with distinct substitutions in common genes. Sci Rep. 2015;5:16550.

48. Li W-H. Molecular evolution. Sunderland, Massachussets: Sinauers Associates; 1997.

49. Stager M, Cerasale DJ, Dor R, Winkler DW, Cheviron ZA. Signatures of natural selection in the mitochondrial genomes of Tachycineta swallows and their implications for latitudinal patterns of the 'pace of life'. Gene. 2014;546(1):104–11.

50. Barrow JR, Stadler HS, Capecchi MR. Roles of Hoxa1 and Hoxa2 in patterning the early hindbrain of the mouse. Development. 2000;127:933–44.

51. Manzanares M, Bel-Vialar S, Aria-McNaughton L, Ferretti E, Marshall H, Maconochie MM, Blasi F, Krumlauf R. Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms. Development. 2001;128:3595–607.

52. Arenkiel BR, Tvrdik P, Gaufo GO, Capecchi MR. Hoxb1 functions in both motoneurons and in tissues of the periphery to establish and maintain the proper neuronal circuitry. Genes Dev. 2004;18(13):1539–52.

53. McNulty CL, Peres JN, Bardine N, van den Akker WM, Durston AJ. Knockdown of the complete Hox paralogous group 1 leads to dramatic hindbrain and neural crest defects. Development. 2005;132(12):2861–71.

54. Duboule D. The rise and fall of Hox gene clusters. Development. 2007;134:2549–60.

55. Ramfrez-Solis R, Zheng H, Whiting J, Krumlauf R, Bradley A. Hoxb-4 (Hox-2.6) mutant mice show homeotic transformation of a cervical vertebra and defects in the closure of the sternal rudiments. Cell. 1993;73(2):279–94.

56. Horan GS, Ramírez-Solis R, Featherstone MS, Wolgemuth DJ, Bradley A, Behringer RR. Compound mutants for the paralogous hoxa-4, hox-b4, and hox-d4 genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed. Genes Dev. 1995;9:1667–77.

57. Buchholtz EA. Modular evolution of the Cetacean vertebral column. Evol Dev. 2007;9(3):278–89.

58. Deschamps J. Developmental biology. Hox genes in the limb: a play in two acts. Science. 2004;304(5677):1610–1.

59. Freitas R, Zhang G, Cohn MJ. Evidence that mechanisms of fin development evolved in the midline of early vertebrates. Nature. 2006;442(7106):1033–7.

60. Delpretti S, Zakany J, Duboule D. A function for all posterior Hoxd genes during digit development? Dev Dyn. 2012;241:792–802.

61. Hostikka SL, Gong J, Carpenter EM. Axial and appendicular skeletal transformations, ligament alterations, and motor neuron loss in Hoxc10 mutants. Int J Biol Sci. 2009;5(5):297–410.

62. Krumlauf R. Hox genes in vertebrate development. Cell. 1994;78:191–201.

63. Soshnikova N, Dewaele R, Janvier P, Krumlauf R, Duboule D. Duplications of hox gene clusters and the emergence of vertebrates. Dev Biol. 2013;378(2):194–9.

64. Wang Z, Yuan L, Rossiter SJ, Zuo X, Ru B, Zhong H, Han N, Jones G, Jepson PD, Zhang S. Adaptive evolution of 5'HoxD Genes in the origin and diversification of the cetacean flipper. Mol Biol Evol. 2009;26(3):613–22.

65. Papageorgiou S. HOX gene expression. New York: Springer Press; 2007.

66. Liang L, Shen Y-Y, Pan X-W, Zhou T-C, Yang C, Irwin DM, Zhang Y-P. Adaptive Evolution of the Hox Gene Family for Development in Bats and Dolphins. PLoS One. 2013;8(6):e65944.

67. Xu B, Wellik DM. Axial Hox9 activity establishes the posterior field in the developing forelimb. Proc Natl Acad Sci U S A. 2011;108(12):4888–91.

68. Dobler S, Dalla S, Wagschal V, Agrawal A. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na, K-ATPase P. Natl Acad Sci USA. 2012;109:13040–5.

69. Projecto-Garcia J, Natarajan C, Moriyama H, Weber R, Fago A, Cheviron Z, et al. Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. Proc Natl Acad Sci. 2013;110:20669–74.

70. Liu Z, Qi F-Y, Zhou X, Ren H-Q, Shi P. Parallel sites implicate functional convergence of the hearing gene prestin among echolocating mammals. Mol Biol Evol. 2014;31:2415–24.

71. Woodard H, Fischman B, Venkat A, Hudson M, Varala K, Cameron S, et al. Genes involved in convergent evolution of eusociality in bees. Proc Natl Acad Sci. 2011;108:7472–7.

72. Pfenninger M, Patel S, Arias-Rodriguez L, Feldmeyer B, Riesch R, Plath M. Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic habitats of a neotropical fish (*Poecilia mexicana*). Mol Ecol. 2015;24:5446–59.

73. Natarajan C, Projecto-Garcia J, Moriyama H, Weber RE, Muñoz-Fuentes V, Green AJ, et al. Convergent evolution of hemoglobin function in high-altitude andean waterfowl involves limited parallelism at the molecular sequence level. Plos Genetics. 2015;11:e1005681.

74. Woods R, Schneider D, Winkworth C, Riley M, Lenski R. Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. Proc Natl Acad Sci. 2006;103:9107–12.

75. Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, et al. The molecular diversity of adaptive convergence. Science. 2012;335:457–61.

76. McClellan DA, Ellison DD. Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. Int J Bioinform Appl. 2010;6(2):120–33.

77. Hallström BM, Janke A. Resolution among major placental mammal interordinal relationships with genome data imply that speciation influenced their earliest radiations. BMC Evol Biol. 2008;8(1):162.

78. Nery MF, González DJ, Hoffmann FG, Opazo JC. Resolution of the laurasiatherian phylogeny: Evidence from genomic data. Mol Phylog Evol. 2012;64:685–9.