The transcriptome of the bowhead whale *Balaena mysticetus* reveals adaptations of the longest-lived mammal

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation


Citable link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:13581028

Terms of Use

This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
The transcriptome of the bowhead whale *Balaena mysticetus* reveals adaptations of the longest-lived mammal

Inge Seim¹, Siming Ma¹, Xuming Zhou¹, Maxim V. Gerashchenko¹, Sang-Goo Lee¹, Robert Suydam², John C. George², John W. Bickham³, Vadim N. Gladyshev¹

¹ Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
² Department of Wildlife Management, North Slope Borough, Barrow, AK 99723, USA
³ Battelle Memorial Institute, Houston, TX 77079, USA

**Key words**: bowhead whale, transcriptome, insulin/IGF1-axis, aging, evolution

**Received**: 10/21/14; **Accepted**: 11/02/14; **Published**: 11/03/14

**Correspondence to**: Vadim N. Gladyshev, PhD; **E-mail**: vgladyshev@rics.bwh.harvard.edu

**Copyright**: Seim et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Abstract**: Mammals vary dramatically in lifespan, by at least two-orders of magnitude, but the molecular basis for this difference remains largely unknown. The bowhead whale *Balaena mysticetus* is the longest-lived mammal known, with an estimated maximal lifespan in excess of two hundred years. It is also one of the two largest animals and the most cold-adapted baleen whale species. Here, we report the first genome-wide gene expression analyses of the bowhead whale, based on the *de novo* assembly of its transcriptome. Bowhead whale or cetacean-specific changes in gene expression were identified in the liver, kidney and heart, and complemented with analyses of positively selected genes. Changes associated with altered insulin signaling and other gene expression patterns could help explain the remarkable longevity of bowhead whales as well as their adaptation to a lipid-rich diet. The data also reveal parallels in candidate longevity adaptations of the bowhead whale, naked mole rat and Brandt's bat. The bowhead whale transcriptome is a valuable resource for the study of this remarkable animal, including the evolution of longevity and its important correlates such as resistance to cancer and other diseases.

**INTRODUCTION**

The bowhead whale (*Balaena mysticetus*) is a sentinel baleen whale species of the Arctic [1]. It has a suite of adaptations for life in an intensely cold and ice-bound habitat. These include the thickest blubber of any whale, primarily for lipid storage but also for effective thermoregulation, and a unique bow-shaped head that allows creation of breathing holes by breaking thick sea ice from beneath. Whales belong to the order Cetacea, which includes the suborders Mysticeti (baleen whales) and Odontoceti (toothed whales, including dolphins, porpoises, sperm whales, etc.), and is a subgroup nested within Artiodactyla (even-toed ungulates such as the hippopotamus, camels, pig, cow, etc.) [2, 3].

Cetaceans are generally long-lived species and, similarly to humans, display traits such as delayed sexual maturity, low fecundity and high survival rates. In contrast to their terrestrial relatives within the order Artiodactyla, which feed on a carbohydrate-rich diet, cetaceans subsist on a lipid-rich diet of krill and other small marine animals. The bowhead whale is the longest-lived mammal known, with an estimated maximal lifespan up to 211 years [4, 5]. The bowhead’s lifespan far exceeds that of other renowned long-lived species of mammals studied for molecular insights into aging, including the naked mole rat (31 years) (*Heterocephalus glaber*) [6] and Brandt’s bat (41 years) (*Myotis brandtii*) [7]. The genomic sequences of only a few cetaceans, including the minke whale (*Balaenoptera...
acutorostrata) [8], bottlenose dolphin (*Tursiops truncatus*) [9], killer whale (*Orcinus orca*) [10] and Yangtze River dolphin (baiji; *Lipotes vexillifer*) [11], have been determined. However, limited access to tissues of these animals has precluded detailed analyses of biological functions based on gene expression.

As a first step in identifying such patterns, we present the liver, kidney and heart transcriptomes of the bowhead whale. Comparison of the bowhead whale transcriptome with that of the related minke whale and other mammals enabled us to identify candidate genes for the exceptional longevity of the bowhead whale as well as molecular adaptations to a lipid-rich diet. Analyses of gene sequences and expression patterns also informed on various aspects of the biology and evolution of bowhead whales.

**RESULTS AND DISCUSSION**

We sequenced the liver, kidney and heart transcriptomes of bowhead whales by Illumina RNA-seq technology. The tissue samples (Fig. 1 and Supplemental Table 1) were sourced from native Iñupiaq Eskimo subsistence harvests in Barrow, Alaska. Using the *de novo* assembler Trinity [12, 13], approximately 659 million paired-end short reads from the bowhead whale were assembled into a single transcriptome per tissue (Supplemental Table 2). Using the Ensembl mouse gene dataset [14] as reference, we identified 9,395 protein-coding genes with one-to-one orthology between diverse mammalian taxa, including whales, a bat, rodents, a tree shrew and a primate (Fig. 2A).

Phylogenetic analysis of the gene expression dataset revealed that the bowhead and minke whales cluster within one evolutionary branch, with a sister group comprising the cow and yak. The bowhead whale and minke whale diverged approximately 13.3 million years ago (Mya) (Fig. 2B). The estimated molecular divergence time of bowhead and minke whales is slightly less than the estimates from comprehensive morphological and molecular phylogenetic analyses of cetacean phylogeny, examining both extant and fossil lineages in simultaneous analyses [15, 16]. It is thus likely that the common ancestor of the bowhead and minke whales lived approximately 13-27 million years ago.

To reveal candidate genes important in the bowhead’s longevity, we compared the bowhead whale transcriptome with that of the minke whale (maximum lifespan of 50 years in the wild [17]), cow, yak, Brandt's bat, Chinese tree shrew, naked mole rat, rhesus macaque, mouse and rat. We also identified unique amino acid changes and rapidly evolving genes in the bowhead whale that may contribute to slow aging in this species.

![Figure 1. The bowhead whale and overview of its geographic range. (A) Bowhead whale (*Balaena mysticetus*). Reproduced courtesy of Encyclopaedia Britannica, Inc., copyright 2002. Human and elephant are shown for size comparison. (B) Occurrence of *B. mysticetus* in Arctic waters. The site of sample collection (Barrow, Alaska) is shown as a red dot.](image-url)
Liver

The liver performs several functions vital to whole-body homeostasis. The integrity of liver function is thus important to longevity, and this organ has gained considerable attention in aging research. A total of 45 genes were differentially expressed in the bowhead whale liver compared to other mammals (Fig. 3, and Tables 1 and 2). In particular, the bowhead whale showed greatly reduced expression of growth factor receptor-bound protein 14 (Grb14) (Fig. 4A). Grb14 is normally highly expressed in the liver and kidney of rodents and humans [18]. In contrast to the liver, Grb14 showed moderate expression in the kidney of the bowhead whale (data not shown). GRB14 binds the receptors for insulin (IR) and IGF1 (IGF1R) to modulate downstream signaling [19-22]. Insulin itself decreases endogenous glucose production (gluconeogenesis) and promotes lipogenesis. Knockout of Grb14 in mouse reduces the amount of circulating insulin and enhances insulin responsiveness in the liver [21]. Similar effects were observed in primary murine hepatocytes, where RNAi knockdown of Grb14 augmented insulin-activated signaling and associated decreased expression of genes associated with gluconeogenesis [23]. During fasting and calorie restriction hepatic gluconeogenesis is required for tissues that lack enzymes required to metabolize free fatty acids (e.g. the brain) [24]. Elevated expression of Cited2 was observed in the bowhead whale (Fig. 4B).

CITED2 is upregulated during fasting, increases the activity of transcriptional co-activator PGC-1α, and directly induces hepatic gluconeogenesis [25]. Interestingly, a recent study found that elevated PGC-1α activity retards aging in Drosophila (26) and it has been hypothesized to also influence longevity in mammals (27). These data suggest that gluconeogenesis is sustained in the bowhead whale despite reduced expression of Grb14. Surprisingly, knockdown of Grb14 in mouse hepatocytes also potently inhibits maturation of sterol regulatory element binding transcription factor 1 (Srebf1, often denoted as SREBP1) and decreases lipogenesis, rendering lipogenic genes unresponsive to insulin [23]. In agreement, we observed decreased expression of insulin-induced gene 1 (Insig1, a downstream target of SREBP1 [23, 28-29]) in the bowhead whale liver (Fig. 4C), indicative of reduced activities along the GRB14/SREBP1 axis.

Figure 2. Overview of orthologous genes shared between the bowhead whale and other mammals and the phylogenetic position of the bowhead whale. (A) Venn diagram showing the number of common and unique orthologs identified in the liver, kidney and heart. (B) Estimation of divergence time using phylogenetic analysis of 10 species based on 6,998 orthologous genes from the liver. Triangle arrows and the numbers above denote the most recent common ancestor (MRCA), and the scale units are million years ago. The standard error range for each divergence time is represented by a dashed line. The red solid circles on the branch nodes denote the node as an ‘age constraint’ used in the estimation of the time of divergence.
Table 1. Genes differentially expressed in the liver of bowhead whales compared to other mammals.

For each gene, gene counts were normalized across all replicates. We used an absolute value of log2 Ratio≥ 2, a Benjamini-Hochberg corrected P-values 0.05, and a B-value of at least 2.945 (representing a 95% probability that a gene is differentially expressed) as the threshold to judge the significance of gene expression difference between the bowhead whale and other mammals. A negative fold change denotes a higher gene expression compared to the other mammals examined, and vice versa.

<table>
<thead>
<tr>
<th>Gene symbol (mouse)</th>
<th>Ensembl transcript ID</th>
<th>Name</th>
<th>logFC</th>
<th>P-value</th>
<th>BH-adj. P-value</th>
<th>B-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams13</td>
<td>ENSMUST00000102891</td>
<td>a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif</td>
<td>-3.84</td>
<td>5.55E-006</td>
<td>8.95E-004</td>
<td>4.06</td>
</tr>
<tr>
<td>Arap3</td>
<td>ENSMUST0000042944</td>
<td>ArfGAP with RhoGAP domain</td>
<td>-2.40</td>
<td>9.01E-007</td>
<td>1.96E-004</td>
<td>5.81</td>
</tr>
<tr>
<td>Atp6v1b2</td>
<td>ENSMUST0000006435</td>
<td>ATPase</td>
<td>2.94</td>
<td>1.42E-008</td>
<td>1.34E-005</td>
<td>9.66</td>
</tr>
<tr>
<td>BC027231</td>
<td>ENSMUST0000048788</td>
<td>cDNA sequence BC027231</td>
<td>-3.12</td>
<td>6.29E-007</td>
<td>1.61E-004</td>
<td>6.15</td>
</tr>
<tr>
<td>Bgn</td>
<td>ENSMUST0000033741</td>
<td>biglycan</td>
<td>-2.45</td>
<td>8.00E-006</td>
<td>1.13E-003</td>
<td>3.70</td>
</tr>
<tr>
<td>Card10</td>
<td>ENSMUST00000164826</td>
<td>caspase recruitment domain family</td>
<td>-3.61</td>
<td>2.61E-007</td>
<td>8.67E-005</td>
<td>7.00</td>
</tr>
<tr>
<td>Cited2</td>
<td>ENSMUST0000038107</td>
<td>Cbp/p300-interacting transactivator</td>
<td>-3.08</td>
<td>1.14E-005</td>
<td>1.37E-003</td>
<td>3.37</td>
</tr>
<tr>
<td>Cxx1b</td>
<td>ENSMUST0000088778</td>
<td>CAAX box 1B</td>
<td>-3.87</td>
<td>3.26E-007</td>
<td>9.67E-005</td>
<td>6.78</td>
</tr>
<tr>
<td>Cyp2r1</td>
<td>ENSMUST0000032908</td>
<td>cytochrome P450</td>
<td>3.42</td>
<td>5.80E-007</td>
<td>1.62E-004</td>
<td>6.21</td>
</tr>
<tr>
<td>Dld1</td>
<td>ENSMUST0000028917</td>
<td>D-tyrosyl-tRNA deacylase 1</td>
<td>-2.16</td>
<td>6.96E-007</td>
<td>1.64E-004</td>
<td>6.06</td>
</tr>
<tr>
<td>Epg5</td>
<td>ENSMUST0000044622</td>
<td>ectopic P-granules autophagy protein 5 homolog (C. elegans)</td>
<td>-4.10</td>
<td>2.70E-006</td>
<td>5.26E-004</td>
<td>4.76</td>
</tr>
<tr>
<td>Epha2</td>
<td>ENSMUST0000006614</td>
<td>Eph receptor A2</td>
<td>-2.76</td>
<td>5.78E-006</td>
<td>8.98E-004</td>
<td>4.03</td>
</tr>
<tr>
<td>Fam129b</td>
<td>ENSMUST0000028135</td>
<td>family with sequence similarity 129</td>
<td>-2.97</td>
<td>3.10E-007</td>
<td>9.67E-005</td>
<td>6.83</td>
</tr>
<tr>
<td>Fam92a</td>
<td>ENSMUST0000108285</td>
<td>family with sequence similarity 92</td>
<td>2.96</td>
<td>6.59E-007</td>
<td>1.62E-004</td>
<td>6.09</td>
</tr>
<tr>
<td>Fbln5</td>
<td>ENSMUST0000021603</td>
<td>fibulin 5</td>
<td>-4.57</td>
<td>5.76E-008</td>
<td>4.65E-005</td>
<td>8.43</td>
</tr>
<tr>
<td>Grb14</td>
<td>ENSMUST0000028252</td>
<td>growth factor receptor bound protein 14</td>
<td>5.21</td>
<td>4.61E-009</td>
<td>5.20E-006</td>
<td>10.73</td>
</tr>
<tr>
<td>Ifi122</td>
<td>ENSMUST0000112923</td>
<td>intraflagellar transport 122</td>
<td>-3.02</td>
<td>6.41E-006</td>
<td>9.52E-004</td>
<td>3.93</td>
</tr>
<tr>
<td>Igi</td>
<td>ENSMUST0000087033</td>
<td>immunoglobulin joining chain</td>
<td>-3.30</td>
<td>9.15E-006</td>
<td>1.20E-003</td>
<td>3.59</td>
</tr>
<tr>
<td>Gene</td>
<td>Accession</td>
<td>Description</td>
<td>Fold Change</td>
<td>p-Value</td>
<td>q-Value</td>
<td>Z-Score</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>Insig1</td>
<td>ENSMUST0000059155</td>
<td>insulin induced gene 1</td>
<td>3.04</td>
<td>8.94E-006</td>
<td>1.20E-003</td>
<td>3.63</td>
</tr>
<tr>
<td>Ints2</td>
<td>ENSMUST0000018212</td>
<td>integrator complex subunit 2</td>
<td>-2.85</td>
<td>5.89E-006</td>
<td>8.98E-004</td>
<td>4.01</td>
</tr>
<tr>
<td>Iqsec1</td>
<td>ENSMUST0000189881</td>
<td>IQ motif and Sec7 domain 1</td>
<td>5.16</td>
<td>2.00E-007</td>
<td>7.52E-005</td>
<td>7.22</td>
</tr>
<tr>
<td>Lgals3</td>
<td>ENSMUST0000142734</td>
<td>lectin</td>
<td>-3.58</td>
<td>1.15E-007</td>
<td>6.48E-005</td>
<td>7.78</td>
</tr>
<tr>
<td>Lrrfip2</td>
<td>ENSMUST0000035078</td>
<td>leucine rich repeat (in FLII) interacting protein 2</td>
<td>2.61</td>
<td>2.52E-007</td>
<td>8.67E-005</td>
<td>6.99</td>
</tr>
<tr>
<td>Masp2</td>
<td>ENSMUST0000052060</td>
<td>mannan-binding lectin serine peptidase 2</td>
<td>3.46</td>
<td>2.63E-009</td>
<td>4.94E-006</td>
<td>11.32</td>
</tr>
<tr>
<td>Med28</td>
<td>ENSMUST0000156481</td>
<td>mediator complex subunit 28</td>
<td>3.16</td>
<td>1.93E-006</td>
<td>3.89E-004</td>
<td>5.08</td>
</tr>
<tr>
<td>Nemf</td>
<td>ENSMUST0000021368</td>
<td>nuclear export mediator factor</td>
<td>3.48</td>
<td>3.17E-006</td>
<td>5.77E-004</td>
<td>4.61</td>
</tr>
<tr>
<td>Ngef</td>
<td>ENSMUST0000068681</td>
<td>neuronal guanine nucleotide exchange factor</td>
<td>5.16</td>
<td>1.26E-007</td>
<td>6.48E-005</td>
<td>7.65</td>
</tr>
<tr>
<td>Nos3</td>
<td>ENSMUST0000030834</td>
<td>nitric oxide synthase 3</td>
<td>-3.58</td>
<td>7.94E-006</td>
<td>1.13E-003</td>
<td>3.72</td>
</tr>
<tr>
<td>Orc5</td>
<td>ENSMUST0000030872</td>
<td>origin recognition complex</td>
<td>2.69</td>
<td>1.92E-006</td>
<td>3.89E-004</td>
<td>5.09</td>
</tr>
<tr>
<td>Pdgfrb</td>
<td>ENSMUST0000115274</td>
<td>platelet derived growth factor receptor</td>
<td>-4.44</td>
<td>3.62E-006</td>
<td>6.38E-004</td>
<td>4.48</td>
</tr>
<tr>
<td>Perp</td>
<td>ENSMUST0000019998</td>
<td>PERP</td>
<td>4.91</td>
<td>6.10E-010</td>
<td>3.44E-006</td>
<td>12.57</td>
</tr>
<tr>
<td>Pglyrp2</td>
<td>ENSMUST0000170392</td>
<td>peptidoglycan recognition protein 2</td>
<td>6.23</td>
<td>1.25E-009</td>
<td>3.52E-006</td>
<td>11.93</td>
</tr>
<tr>
<td>Rasip1</td>
<td>ENSMUST0000057927</td>
<td>Ras interacting protein 1</td>
<td>-4.69</td>
<td>9.75E-008</td>
<td>6.11E-005</td>
<td>7.94</td>
</tr>
<tr>
<td>Rnase4</td>
<td>ENSMUST0000022428</td>
<td>ribonuclease</td>
<td>3.21</td>
<td>3.60E-007</td>
<td>1.01E-004</td>
<td>6.69</td>
</tr>
<tr>
<td>Rpa2</td>
<td>ENSMUST00000102561</td>
<td>replication protein A2</td>
<td>-2.42</td>
<td>2.87E-006</td>
<td>5.41E-004</td>
<td>4.70</td>
</tr>
<tr>
<td>Rpusd4</td>
<td>ENSMUST0000034543</td>
<td>RNA pseudouridylate synthase domain containing 4</td>
<td>-4.13</td>
<td>1.84E-007</td>
<td>7.52E-005</td>
<td>7.33</td>
</tr>
<tr>
<td>Scl25a38</td>
<td>ENSMUST0000035106</td>
<td>solute carrier family 25</td>
<td>2.90</td>
<td>8.36E-007</td>
<td>1.89E-004</td>
<td>5.87</td>
</tr>
<tr>
<td>Smad2</td>
<td>ENSMUST0000168423</td>
<td>SMAD family member 2</td>
<td>2.12</td>
<td>1.58E-005</td>
<td>1.81E-003</td>
<td>3.09</td>
</tr>
<tr>
<td>Sshp4</td>
<td>ENSMUST0000049908</td>
<td>single stranded DNA binding protein 4</td>
<td>-2.50</td>
<td>8.39E-006</td>
<td>1.15E-003</td>
<td>3.68</td>
</tr>
<tr>
<td>Tie1</td>
<td>ENSMUST0000047421</td>
<td>tyrosine kinase with immunoglobulin-like and EGF-like domains 1</td>
<td>-3.79</td>
<td>8.04E-008</td>
<td>5.67E-005</td>
<td>8.12</td>
</tr>
<tr>
<td>Gene</td>
<td>Identifier</td>
<td>Description</td>
<td>Fold Change</td>
<td>Benjamini–Hochberg FDR</td>
<td>p-value</td>
<td>FDR</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>--------------------------------------------------</td>
<td>-------------</td>
<td>------------------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Timp2</td>
<td>ENSMUST0000017610</td>
<td>tissue inhibitor of metalloproteinase 2</td>
<td>-5.09</td>
<td>3.71E-009</td>
<td>5.20E-006</td>
<td>11.02</td>
</tr>
<tr>
<td>Top1</td>
<td>ENSMUST00000109468</td>
<td>topoisomerase (DNA) I</td>
<td>2.18</td>
<td>1.07E-005</td>
<td>1.35E-003</td>
<td>3.45</td>
</tr>
<tr>
<td>Vac14</td>
<td>ENSMUST0000034190</td>
<td>Vac14 homolog (S. cerevisiae)</td>
<td>-2.27</td>
<td>1.87E-007</td>
<td>7.52E-005</td>
<td>7.32</td>
</tr>
<tr>
<td>Wsb1</td>
<td>ENSMUST0000017821</td>
<td>WD repeat and SOCS box-containing 1</td>
<td>-4.40</td>
<td>1.57E-007</td>
<td>7.40E-005</td>
<td>7.48</td>
</tr>
<tr>
<td>Zfp143</td>
<td>ENSMUST0000084727</td>
<td>zinc finger protein 143</td>
<td>-2.43</td>
<td>5.19E-006</td>
<td>8.62E-004</td>
<td>4.13</td>
</tr>
</tbody>
</table>

**Figure 3.** Heat map displaying 45 genes differentially expressed in the bowhead whale liver. Genes with ≥2 log2-fold change and with a Benjamini–Hochberg false discovery rate ≤ 0.05 are shown. The color scheme was used wherein the gradient of red indicates decreased expression, and blue increased expression. Selected genes discussed in the text are highlighted. Hierarchical clustering of samples was based on Pearson’s correlation of gene expression data.
Calorie restriction and modulation of the insulin/IGF1-signaling pathway are key modulators of aging and have potent effects on lipid homeostasis [30]. Lipid metabolism, therefore, plays a central role in lifespan regulation in metazoan eukaryotes. While carbohydrates are a major contributor to metabolism of terrestrial mammals, marine mammals subsist on a high-energy, lipid-rich diet [31]. Notably, the bowhead whale diet consists of zooplankton with very high lipid content [32, 33]. Toothed whales, such as the common bottlenose dolphin, which has a maximum lifespan of ~52 years in captivity [34], can exhibit pathology that parallels type two diabetes (T2D) and metabolic syndrome of humans. This includes a high prevalence of insulin resistance, fatty liver disease, and chronic overproduction of Very-low-density lipoprotein (VLDL) by the liver (dyslipidemia) [35-37]. Our data suggest that reduced \textit{Grb14} expression in the bowhead whale improves energy homeostasis in an animal that depends on an exceptionally lipid-rich diet. This beneficial adaptation may serve to protect the long-lived bowhead whale against chronic dietary diseases.

Figure 4. Boxplots of selected genes differentially expressed in the bowhead whale liver. (A) Growth factor receptor bound protein 14 (\textit{Grb14}). (B) Insulin induced gene 1 (\textit{Insig1}). (C) Cbp/p300-interacting transactivator (\textit{Cited2}). (D) TP53 Apoptosis Effector (\textit{Perp}). Genes were subjected to limma analysis at 5% FDR and 2.0 fold-change cut-off. Normalized log2 read counts and P-values adjusted for multiple comparisons (Benjamini and Hochberg) are shown.
It has been proposed that the longevity of large mammals, such as the bowhead whale and elephant, stem from a lack of non-human predators, slow development and concomitant tight control of nutrient sensing pathways, in particular the insulin/TOR signaling axis [38]. Small African mole rats and many bats live in protected environments and are long-lived relative to animals of similar size, but may in contrast have evolved genetic changes that allow them to slow down aging after a comparatively rapid development [38]. Sequence and expression changes of genes involved in hepatic metabolism have been reported for Brandt's bat [39] and naked mole rat [40], two terrestrial species in our dataset that are long-lived compared to

<table>
<thead>
<tr>
<th>GO Category</th>
<th>Term</th>
<th>Count</th>
<th>Genes</th>
<th>Fold Enrichment</th>
<th>Fisher's exact P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0001501</td>
<td>skeletal system development</td>
<td>4</td>
<td>Epha2, Insig1, Lgals3, Pdgfrb</td>
<td>7.3</td>
<td>2.00E-003</td>
</tr>
<tr>
<td>GO:0048568</td>
<td>embryonic organ development</td>
<td>3</td>
<td>Epha2, Cited2, Insig1</td>
<td>7.4</td>
<td>7.20E-003</td>
</tr>
<tr>
<td>GO:0006260</td>
<td>DNA replication</td>
<td>3</td>
<td>Orc5, Rpa2, Top1</td>
<td>6.1</td>
<td>1.30E-002</td>
</tr>
<tr>
<td>GO:0040013</td>
<td>negative regulation of locomotion</td>
<td>2</td>
<td>Arap3, Tie1</td>
<td>22.1</td>
<td>3.50E-003</td>
</tr>
<tr>
<td>GO:0030336</td>
<td>negative regulation of cell migration</td>
<td>2</td>
<td>Arap3, Tie1</td>
<td>22.1</td>
<td>3.50E-003</td>
</tr>
<tr>
<td>GO:0001570</td>
<td>vasculogenesis</td>
<td>2</td>
<td>Epha2, Cited2</td>
<td>20.6</td>
<td>4.00E-003</td>
</tr>
<tr>
<td>GO:0051271</td>
<td>negative regulation of cell motion</td>
<td>2</td>
<td>Arap3, Tie1</td>
<td>19.2</td>
<td>4.60E-003</td>
</tr>
</tbody>
</table>

**Table 2. Enrichment of biological process (BP) Gene Ontology (GO) terms for genes differentially expressed in the bowhead liver compared to other mammals.**

**Table 3. Genes differentially expressed in the bowhead and minke whale heart compared to other mammals.**

For each gene, gene counts were normalized across all replicates. We used an absolute value of log2 Ratios2, a Benjamini-Hochberg corrected P-values 0.05, and a B-value of at least 2.945 (representing a 95% probability that a gene is differentially expressed) as the threshold to judge the significance of gene expression difference between whales and other mammals. A negative fold change denotes a higher gene expression compared to the other mammals examined, and vice versa.

<table>
<thead>
<tr>
<th>Gene symbol (mouse)</th>
<th>Ensembl transcript ID</th>
<th>Name</th>
<th>logFC</th>
<th>P-value</th>
<th>BH-adj. P-value</th>
<th>B-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asl</td>
<td>ENSMUST00000161094</td>
<td>argininosuccinate lyase</td>
<td>-4.40</td>
<td>1.43E-005</td>
<td>1.50E-002</td>
<td>3.19</td>
</tr>
<tr>
<td>Brk1</td>
<td>ENSMUST0000035725</td>
<td>BRICK1, SCAR/WAVE actin-nucleating complex subunit</td>
<td>-2.85</td>
<td>5.69E-006</td>
<td>7.93E-003</td>
<td>3.96</td>
</tr>
<tr>
<td>Eif3l</td>
<td>ENSMUST0000040518</td>
<td>eukaryotic translation initiation factor 3, subunit L</td>
<td>-2.46</td>
<td>2.28E-006</td>
<td>4.77E-003</td>
<td>4.95</td>
</tr>
<tr>
<td>Hiatip2</td>
<td>ENSMUST0000085272</td>
<td>HIV-1 tat interactive protein 2, homolog (human)</td>
<td>-6.00</td>
<td>1.40E-007</td>
<td>5.87E-004</td>
<td>6.99</td>
</tr>
</tbody>
</table>
related species. Interestingly, in agreement with the previous studies that compared a smaller set of species, we found that Foxo1 (forkhead box O1) and Fto (fat mass and obesity associated) were differentially expressed in liver of the Brandt's bat and naked mole rat, respectively (Supplemental Tables 3-6). The transcription factor FOXO1, like PGC-1α, is a classical mediator of gluconeogenesis in species ranging from worms to mammals, and Foxo1 and its upstream regulator Creb1 [24] are expressed at a higher level in the Brandt's bat liver (Supplemental Table 3). Elevated hepatic Foxo1 mRNA expression in the Brandt's bat reconciles with increased longevity in calorie restricted and growth hormone receptor knockout mice [41], the worm C. elegans [42] and the fruit fly Drosophila melanogaster [43]. This result strongly suggests that elevated hepatic Foxo1 expression contributes to the longevity of the Brandt's bat. The naked mole rat has a highly divergent insulin peptide that may be compensated by autocrine/paracrine IGF2 in the adult liver and displays physiological changes consistent with an altered insulin/IGF1 axis [40, 44]. In the naked mole rat, fat mass and obesity associated mRNA (encoded by Fto) is expressed at a lower level compared to the other species examined (Supplemental Table 5). Fto regulates glucose metabolism in the liver and its reduced expression protects against obesity and metabolic syndrome, and associated pathology such as insulin resistance [45-47]. Taken together, we propose that fundamental longevity-promoting mechanisms in the long-lived bowhead whale, Brandt's bat and naked mole rat stem from overlapping but distinct expression and sequence changes in metabolism genes.

Table 4. Proteins with unique amino acids in the bowhead whale compared to 67 vertebrates.

AA changes correspond to residue locations in the translated human RefSeq mRNA entry (typically the longest RefSeq entry of each gene). Genes shown in bold were checked by BLAST analysis of raw sequence data. Some of the genes in this table, especially those not shown in bold, may be false-positives.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Symbol (human)</th>
<th>RefSeq ID</th>
<th>AA change(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>actin-binding Rho activating protein</td>
<td>ABRA</td>
<td>NM_139166</td>
<td>HKR283P</td>
</tr>
<tr>
<td>adaptor-related protein complex 3, beta 1 subunit</td>
<td>AP3B1</td>
<td>NM_003664</td>
<td>R1086Q</td>
</tr>
<tr>
<td>annexin A7</td>
<td>ANXA7</td>
<td>NM_001156</td>
<td>Y410C</td>
</tr>
<tr>
<td>arachidonate 5-lipoxygenase</td>
<td>ALOX5</td>
<td>NM_000698</td>
<td>D204N</td>
</tr>
<tr>
<td>calcium homeostasis modulator 2</td>
<td>CALHM2</td>
<td>NM_015916</td>
<td>F172C</td>
</tr>
<tr>
<td>calsyntenin 2</td>
<td>CLSTN2</td>
<td>NM_022131</td>
<td>Q362H</td>
</tr>
<tr>
<td>cell division cycle associated 7</td>
<td>CDCA7</td>
<td>NM_145810</td>
<td>ED257K</td>
</tr>
<tr>
<td>c-fos induced growth factor (vascular endothelial growth factor D)</td>
<td>FIGF</td>
<td>NM_004469</td>
<td>STQ174R</td>
</tr>
<tr>
<td>ClpB caseinolytic peptidase B homolog (E. coli)</td>
<td>CLPB</td>
<td>NM_001258392</td>
<td>HR226W</td>
</tr>
<tr>
<td>complement component 5</td>
<td>C5</td>
<td>NM_001735</td>
<td>I760K</td>
</tr>
<tr>
<td>cytidine and dCMP deaminase domain containing 1</td>
<td>CDADC1</td>
<td>NM_030911</td>
<td>YQ367C</td>
</tr>
<tr>
<td>derlin 1</td>
<td>DERL1</td>
<td>NM_024295</td>
<td>R197G</td>
</tr>
<tr>
<td>enhancer of polycomb homolog 1 (Drosophila)</td>
<td>EPC1</td>
<td>NM_001272004</td>
<td>G562R</td>
</tr>
<tr>
<td>erythrocyte membrane protein band 4.1-like 1</td>
<td>EPB41L1</td>
<td>NM_001258329</td>
<td>K478M</td>
</tr>
<tr>
<td>fucose-1-phosphate guanylytransferase</td>
<td>FPGT</td>
<td>NM_003838</td>
<td>HK339E</td>
</tr>
<tr>
<td>gem (nucleus organelle) associated protein 4</td>
<td>GEMIN4</td>
<td>NM_015721</td>
<td>KR721T</td>
</tr>
<tr>
<td>gigaxonin</td>
<td>GAN</td>
<td>NM_022041</td>
<td>E74V</td>
</tr>
<tr>
<td>glutamate receptor, ionotropic, kainate 2</td>
<td>GRIK2</td>
<td>NM_175768</td>
<td>Y571C, Q621R</td>
</tr>
<tr>
<td>glutamate receptor, metabotropic 1</td>
<td>GRM1</td>
<td>NM_001278066</td>
<td>E559D</td>
</tr>
</tbody>
</table>
Living under water also entails a unique set of challenges. Compared with the other mammals, the bowhead liver had lower expression of the apoptosis-associated p53 target gene *Perp* (TP53 Apoptosis Effector) (Fig. 4D). Under hypoxic stress, p53 is activated to trigger cellular apoptosis as a protective response, but this may not be appropriate for mammals with only intermittent access to oxygen. It has been shown in mouse kidney that *Perp* knockdown protected cells from hypoxia-induced apoptosis [48]. Low levels of *Perp* may therefore increase stress tolerance in the bowhead whale liver. In addition, we observed increased expression of genes that enhance vasculature maintenance (EPH receptor A2 (*Epha2*) [49]), regulate

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Symbol</th>
<th>Accession</th>
<th>HGNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>heat shock protein 90kDa alpha (cytosolic), class B member 1</td>
<td>HSP90AB1</td>
<td>NM_001271971</td>
<td>H121T</td>
</tr>
<tr>
<td>homeodomain interacting protein kinase 1</td>
<td>HIPK1</td>
<td>NM_181358</td>
<td>SP46SC</td>
</tr>
<tr>
<td>homeodomain interacting protein kinase 3</td>
<td>HIPK3</td>
<td>NM_005734</td>
<td>Q698H</td>
</tr>
<tr>
<td>ligase III, DNA, ATP-dependent</td>
<td>LIG3</td>
<td>NM_013975</td>
<td>D255N</td>
</tr>
<tr>
<td>mitogen-activated protein kinase kinase 7</td>
<td>MAP3K7</td>
<td>NM_145331</td>
<td>V424D</td>
</tr>
<tr>
<td>MMS19 nucleotide excision repair homolog (S. cerevisiae)</td>
<td>MMS19</td>
<td>NM_022362</td>
<td>SLT202C</td>
</tr>
<tr>
<td>Nance-Horan syndrome (congenital cataracts and dental anomalies)</td>
<td>NHS</td>
<td>NM_001291868</td>
<td>R388Q</td>
</tr>
<tr>
<td>NEDD4 binding protein 3</td>
<td>N4BP3</td>
<td>NM_015111</td>
<td>Q384H</td>
</tr>
<tr>
<td>neuronal pentraxin II</td>
<td>NPTX2</td>
<td>NM_002523</td>
<td>Q375H</td>
</tr>
<tr>
<td>NOP2 nucleolar protein</td>
<td>NOP2</td>
<td>NM_001258310</td>
<td>YF516C</td>
</tr>
<tr>
<td>NRDE-2, necessary for RNA interference, domain containing</td>
<td>NRDE2</td>
<td>NM_017970</td>
<td>H592Q</td>
</tr>
<tr>
<td>nuclear receptor coactivator 2</td>
<td>NCOA2</td>
<td>NM_006540</td>
<td>F441C</td>
</tr>
<tr>
<td>nuclear receptor coactivator 7</td>
<td>NCOA7</td>
<td>NM_181782</td>
<td>SAT160D</td>
</tr>
<tr>
<td>nucleolar protein 6 (RNA-associated)</td>
<td>NOL6</td>
<td>NM_022917</td>
<td>R582L</td>
</tr>
<tr>
<td>oncprotein induced transcript 3</td>
<td>OIT3</td>
<td>NM_152635</td>
<td>P353R</td>
</tr>
<tr>
<td>oxidative stress induced growth inhibitor family member 2</td>
<td>OSGIN2</td>
<td>NM_004337</td>
<td>ED194K</td>
</tr>
<tr>
<td>PALM2-AKAP2 readthrough</td>
<td>PALM2- AKAP2</td>
<td>NM_147150</td>
<td>ED488G</td>
</tr>
<tr>
<td>patched 1</td>
<td>PTCH1</td>
<td>NM_000264</td>
<td>S1185C</td>
</tr>
<tr>
<td>phosphodiesterase 4D interacting protein</td>
<td>PDE4DIP</td>
<td>NM_001002811</td>
<td>SGNTP264R</td>
</tr>
<tr>
<td>polymerase (DNA-directed), delta interacting protein 3</td>
<td>POLDIP3</td>
<td>NM_032311</td>
<td>Q198R</td>
</tr>
<tr>
<td>potassium voltage-gated channel, Shal-related subfamily, member 2</td>
<td>KCND2</td>
<td>NM_012281</td>
<td>L457R</td>
</tr>
<tr>
<td>protein kinase, cAMP-dependent, regulatory, type 1, alpha</td>
<td>PRKAR1A</td>
<td>NM_212472</td>
<td>N26D</td>
</tr>
<tr>
<td>protocadherin alpha subfamily C, 2</td>
<td>PCDHAC2</td>
<td>NM_018899</td>
<td>K884Q</td>
</tr>
<tr>
<td>serpin peptidase inhibitor, clade C (antithrombin), member 1</td>
<td>SERPINC1</td>
<td>NM_000488</td>
<td>E359D</td>
</tr>
<tr>
<td>striatin, calmodulin binding protein</td>
<td>STRN</td>
<td>NM_003162</td>
<td>ED279V</td>
</tr>
<tr>
<td>transmembrane protein 214</td>
<td>TMEM214</td>
<td>NM_017727</td>
<td>R251Q</td>
</tr>
<tr>
<td>transmembrane protein 25</td>
<td>TMEM25</td>
<td>NM_032780</td>
<td>G162D</td>
</tr>
<tr>
<td>tripartite motif containing 29</td>
<td>TRIM29</td>
<td>NM_012101</td>
<td>Q250H</td>
</tr>
<tr>
<td>tyrosine aminotransferase</td>
<td>TAT</td>
<td>NM_000353</td>
<td>Y442H</td>
</tr>
<tr>
<td>vacuolar protein sorting 13 homolog B (yeast)</td>
<td>VPS13B</td>
<td>NM_015243</td>
<td>G567R</td>
</tr>
</tbody>
</table>
vascular tone and proliferation (endothelial nitric oxide synthase (eNOS) (Nos3) [50, 51]) and promote DNA repair (replication protein A2 (Rpa2) [52, 53]) (Fig. 3, Tables 1 and 2), which may also represent adaptations to limited oxygen availability and water pressure in the aquatic environment.

**Cardiovascular system**

Other extraordinarily long-lived species, such as the naked mole rat, possess biological adaptations endowing resilience to aging-related diseases, including cardiovascular disease and cancer [54, 55]. The availability of a single heart sample precluded differential gene expression analysis of the bowhead whale alone. Using transcriptome data from a bowhead whale heart and a minke whale heart [8], we compared these marine mammals to the terrestrial ones and identified four differentially expressed genes in whales (Table 3). Of these, argininosuccinate lyase (Asl) is particularly interesting. The expression of Asl, which is essential for NO production in the heart [56], was 4.4-fold higher in the whales (Table 3). Reduction of nitrite to nitric oxide (NO) provides cytoprotection of tissues during hypoxic events [57] and serves to preserve cardiac function in the diving and hibernating red-eared slider turtle (Trachemys scripta elegans) [58]. Cardiovascular regulation is critical during diving in all marine mammal species, and evidence suggests that hypoxia-sensitive tissues such as the brain and heart have adapted to low oxygen conditions while diving [59]. We speculate that elevated expression of Asl improves cardiac metabolism and function of diving cetaceans.

Amino-acid residues that are uniquely altered in a lineage can reveal clues to functional adaptations [39, 60]. In addition to differential gene expression patterns, we identified a unique amino acid change with potential relevance to vascular aging in the bowhead whale (Table 4). An examination of 68 vertebrate genomes, including 6 cetaceans (Supplemental Table 7), revealed a unique amino acid change in bowhead whale c-fos induced growth factor (Figf, coding for vascular endothelial growth factor D, VEGFD) (Fig. 5). VEGFD plays a role in the maintenance of vascular homeostasis [61, 62]. The radical amino acid change, neutral polar residues replaced by a charged residue (SerThrGln174Arg), in bowhead whale VEGFD flanks a large hydrophobic surface that interacts with the cognate receptor VEGFR-2 [63]. Although this finding must be corroborated with experimental data, we speculate that the arginine substitution may serve to improve the interaction of VEGFD with its cognate receptor, potentially aiding maintenance of vascular health in bowhead whales.

**Kidney**

Aging-induced changes in the kidney include reduced repair and/or regeneration of cells, which is concomitant with decreases in glomerular filtration rate and blood flow [64, 65]. Comparison of the bowhead whale kidney transcriptome with those of other mammals revealed that 53 genes were differentially expressed in the bowhead whale kidney (Tables 5 and 6). Among these was a battery of DNA repair associated genes. For example, Sprt-like N-terminal domain (Sprtn) [66]; non-SMC condensin I complex, subunit D2 (Ncapd2) [67]; ligase I, DNA, ATP-dependent (Lig1) [68]; and Rpa2 [52, 53] which also was elevated in the liver, were expressed at higher levels in the bowhead whale kidney (Table 5). The mitochondrial stress inhibitor Jumbl [69], as well as genes associated with tumor suppression, was also expressed at higher levels in the bowhead whale kidney. These latter genes included E4F transcription factor 1 (E4fl) which directly interacts with the tumor suppressors p53, RASSF1A, pRB and p14ARF [70]; ADP-ribosylation factor-like 2 (Arfl2) [71]; protein kinase C, delta binding protein (Prkcdbp; also known as hSRBC) [72, 73]; and Egrf [74]. We speculate that this series of genes protect against age-related kidney tissue decline in the bowhead whale.

Figure 5. Evolution of vertebrate VEGFD protein sequences. (A) Alignment of vertebrate VEGFD protein sequences. VEGFD residues at the VEGFR-2 receptor-ligand interface are highlighted in yellow, highly conserved cysteine residues that stabilize VEGFD in orange, and a radical amino acid in the bowhead whale, but not in 67 other species, is shown in red. (B) Modeled VEGFD structure. Location of the unique arginine residue in the bowhead whale protein is shown.
Expanding on the theme of improved genome integrity and repair, we found a radical amino acid change SerLeuThr202Cys in the HEAT2 domain of bowhead whale MMS19 nucleotide excision repair homolog (encoded by \textit{Mms19}) (Table 4). This highly conserved protein plays an essential role in genome stability by facilitating iron-sulfur (FeS) cluster insertion into proteins involved in methionine biosynthesis, DNA replication and repair and telomere maintenance [75, 76]. Reactive cysteines in assembly complex proteins, such as MMS19, assist in the transfer of FeS clusters to target proteins [77]. The amino acid change in bowhead whale MMS19 may confer systemic robustness to damage and/or longevity.

Table 5. Genes differentially expressed in the kidney of bowhead whales compared to other mammals.

For each gene, gene counts were normalized across all replicates. We used an absolute value of log2 Ratio ≥ 2, Benjamini-Hochberg corrected P-values 0.05 a B-value of at least 2.945 (representing a 95% probability that a gene is differentially expressed) as the threshold to judge the significance of gene expression difference between the bowhead whale and other mammals. A negative fold change denotes a higher gene expression compared to the other mammals examined, and vice versa.

<table>
<thead>
<tr>
<th>Gene symbol (mouse)</th>
<th>Ensembl transcript ID</th>
<th>Name</th>
<th>logFC</th>
<th>P-value</th>
<th>BH-adj. P-value</th>
<th>B-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamts1</td>
<td>ENSMUST0000023610</td>
<td>a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 1</td>
<td>-2.08</td>
<td>2.6E-07</td>
<td>1.0E-04</td>
<td>6.98</td>
</tr>
<tr>
<td>Ankrd6</td>
<td>ENSMUST0000035719</td>
<td>ankyrin repeat domain 6</td>
<td>-3.37</td>
<td>2.5E-06</td>
<td>3.2E-04</td>
<td>4.84</td>
</tr>
<tr>
<td>Arhgef19</td>
<td>ENSMUST000006618</td>
<td>Rho guanine nucleotide exchange factor (GEF) 19</td>
<td>-3.41</td>
<td>2.9E-07</td>
<td>1.0E-04</td>
<td>6.89</td>
</tr>
<tr>
<td>Arl2</td>
<td>ENSMUST0000025893</td>
<td>ADP-ribosylation factor-like 2</td>
<td>-2.36</td>
<td>1.2E-06</td>
<td>2.2E-04</td>
<td>5.52</td>
</tr>
<tr>
<td>Atp5g2</td>
<td>ENSMUST0000185641</td>
<td>ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C2 (subunit 9)</td>
<td>-2.22</td>
<td>8.6E-07</td>
<td>1.8E-04</td>
<td>5.83</td>
</tr>
<tr>
<td>Cdh3</td>
<td>ENSMUST0000080797</td>
<td>cadherin 3</td>
<td>-3.41</td>
<td>7.6E-06</td>
<td>6.6E-04</td>
<td>3.73</td>
</tr>
<tr>
<td>Cep170</td>
<td>ENSMUST0000057037</td>
<td>centrosomal protein 170</td>
<td>-2.71</td>
<td>1.5E-06</td>
<td>2.3E-04</td>
<td>5.36</td>
</tr>
<tr>
<td>Cpsf3l</td>
<td>ENSMUST0000039091</td>
<td>cleavage and polyadenylation specific factor 3-like</td>
<td>-2.44</td>
<td>3.3E-06</td>
<td>3.9E-04</td>
<td>4.52</td>
</tr>
<tr>
<td>Cxx1b</td>
<td>ENSMUST0000088778</td>
<td>CAAX box 1B</td>
<td>-2.29</td>
<td>1.6E-06</td>
<td>2.3E-04</td>
<td>5.25</td>
</tr>
<tr>
<td>Dnnd4b</td>
<td>ENSMUST0000098914</td>
<td>DENN/MADD domain containing 4B</td>
<td>-2.68</td>
<td>2.3E-06</td>
<td>3.0E-04</td>
<td>4.92</td>
</tr>
<tr>
<td>E4f1</td>
<td>ENSMUST0000056032</td>
<td>E4F transcription factor 1</td>
<td>-2.08</td>
<td>1.7E-05</td>
<td>1.1E-03</td>
<td>2.99</td>
</tr>
<tr>
<td>Egr1</td>
<td>ENSMUST0000064795</td>
<td>early growth response 1</td>
<td>-4.04</td>
<td>2.2E-07</td>
<td>9.4E-05</td>
<td>7.16</td>
</tr>
<tr>
<td>Epb4.112</td>
<td>ENSMUST0000053748</td>
<td>erythrocyte protein band 4.1-like 2</td>
<td>-2.41</td>
<td>2.2E-08</td>
<td>2.0E-05</td>
<td>9.36</td>
</tr>
<tr>
<td>Fancg</td>
<td>ENSMUST0000030165</td>
<td>Fanconi anemia, complementation group G</td>
<td>-3.22</td>
<td>1.0E-05</td>
<td>7.6E-04</td>
<td>3.51</td>
</tr>
<tr>
<td>Gene</td>
<td>ENSMUST00000045737</td>
<td>UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 11</td>
<td>2.33</td>
<td>7.9E-06</td>
<td>6.6E-04</td>
<td>3.69</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>------</td>
<td>---------</td>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Gja4</td>
<td>ENSMUST00000053753</td>
<td>gap junction protein, alpha 4</td>
<td>-3.17</td>
<td>7.3E-08</td>
<td>4.8E-05</td>
<td>8.21</td>
</tr>
<tr>
<td>Glscr2</td>
<td>ENSMUST00000044158</td>
<td>glioma tumor suppressor candidate region gene 2</td>
<td>-2.60</td>
<td>2.2E-08</td>
<td>2.0E-05</td>
<td>9.40</td>
</tr>
<tr>
<td>Hdac11</td>
<td>ENSMUST00000041736</td>
<td>histone deacetylase 11</td>
<td>-2.01</td>
<td>8.3E-07</td>
<td>1.8E-04</td>
<td>5.89</td>
</tr>
<tr>
<td>Heatr1</td>
<td>ENSMUST00000059270</td>
<td>HEAT repeat containing 1</td>
<td>2.60</td>
<td>1.2E-06</td>
<td>2.2E-04</td>
<td>5.47</td>
</tr>
<tr>
<td>Hecw2</td>
<td>ENSMUST00000120904</td>
<td>HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2</td>
<td>-2.62</td>
<td>7.8E-07</td>
<td>1.8E-04</td>
<td>5.94</td>
</tr>
<tr>
<td>Hoxa5</td>
<td>ENSMUST00000048794</td>
<td>homeobox A5</td>
<td>-3.44</td>
<td>7.9E-06</td>
<td>6.6E-04</td>
<td>3.73</td>
</tr>
<tr>
<td>Igsf11</td>
<td>ENSMUST00000023478</td>
<td>immunoglobulin superfamily, member 11</td>
<td>4.09</td>
<td>9.2E-07</td>
<td>1.9E-04</td>
<td>5.72</td>
</tr>
<tr>
<td>Junb</td>
<td>ENSMUST00000064922</td>
<td>jun B proto-oncogene</td>
<td>-3.44</td>
<td>1.5E-06</td>
<td>2.3E-04</td>
<td>5.29</td>
</tr>
<tr>
<td>Klf9</td>
<td>ENSMUST00000036884</td>
<td>Kruppel-like factor 9</td>
<td>2.26</td>
<td>9.3E-06</td>
<td>7.2E-04</td>
<td>3.60</td>
</tr>
<tr>
<td>Lig1</td>
<td>ENSMUST00000177588</td>
<td>ligase I, DNA, ATP-dependent</td>
<td>-2.50</td>
<td>9.3E-06</td>
<td>7.2E-04</td>
<td>3.54</td>
</tr>
<tr>
<td>Lrfn4</td>
<td>ENSMUST00000113822</td>
<td>leucine rich repeat and fibronectin type III domain containing 4</td>
<td>-2.78</td>
<td>5.0E-07</td>
<td>1.4E-04</td>
<td>6.37</td>
</tr>
<tr>
<td>Marc2</td>
<td>ENSMUST00000068725</td>
<td>mitochondrial amidoxime reducing component 2</td>
<td>5.01</td>
<td>2.6E-07</td>
<td>1.0E-04</td>
<td>6.90</td>
</tr>
<tr>
<td>Ncapd2</td>
<td>ENSMUST00000043848</td>
<td>non-SMC condensin 1 complex, subunit D2</td>
<td>-2.14</td>
<td>1.3E-08</td>
<td>1.5E-05</td>
<td>9.91</td>
</tr>
<tr>
<td>Ncs1</td>
<td>ENSMUST0000000199</td>
<td>neuronal calcium sensor 1</td>
<td>-2.40</td>
<td>8.7E-07</td>
<td>1.8E-04</td>
<td>5.84</td>
</tr>
<tr>
<td>Nfyα</td>
<td>ENSMUST00000046719</td>
<td>nuclear transcription factor-Y alpha</td>
<td>3.45</td>
<td>2.2E-07</td>
<td>9.4E-05</td>
<td>7.02</td>
</tr>
<tr>
<td>Nipsnap3</td>
<td>ENSMUST00000015391</td>
<td>nipsnap homolog 3B (C. elegans)</td>
<td>-2.47</td>
<td>5.0E-06</td>
<td>5.0E-04</td>
<td>4.18</td>
</tr>
<tr>
<td>Nle1</td>
<td>ENSMUST00000103213</td>
<td>notchless homolog 1 (Drosophila)</td>
<td>-3.32</td>
<td>7.1E-10</td>
<td>2.6E-06</td>
<td>12.57</td>
</tr>
<tr>
<td>Nme6</td>
<td>ENSMUST00000035053</td>
<td>NME/NM23 nucleoside diphosphate kinase 6</td>
<td>2.34</td>
<td>4.7E-07</td>
<td>1.4E-04</td>
<td>6.30</td>
</tr>
<tr>
<td>Nr4a1</td>
<td>ENSMUST00000023779</td>
<td>nuclear receptor subfamily 4, group A, member 1</td>
<td>-5.09</td>
<td>1.5E-07</td>
<td>7.7E-05</td>
<td>7.53</td>
</tr>
<tr>
<td>Pallid</td>
<td>ENSMUST00000121785</td>
<td>palladin, cytoskeletal associated protein</td>
<td>4.13</td>
<td>7.0E-09</td>
<td>1.3E-05</td>
<td>10.13</td>
</tr>
<tr>
<td>Pmf1</td>
<td>ENSMUST00000056370</td>
<td>polyamine-modulated factor 1</td>
<td>-2.39</td>
<td>8.1E-06</td>
<td>6.7E-04</td>
<td>3.72</td>
</tr>
<tr>
<td>Pomk</td>
<td>ENSMUST00000061850</td>
<td>protein-O-mannose kinase</td>
<td>-2.21</td>
<td>1.0E-06</td>
<td>2.0E-04</td>
<td>5.72</td>
</tr>
<tr>
<td>Por</td>
<td>ENSMUST0000005651</td>
<td>P450 (cytochrome) oxidoreductase</td>
<td>2.49</td>
<td>4.8E-06</td>
<td>5.0E-04</td>
<td>4.18</td>
</tr>
<tr>
<td>Prelid1</td>
<td>ENSMUST00000021942</td>
<td>PRELI domain containing 1</td>
<td>2.72</td>
<td>1.4E-07</td>
<td>7.7E-05</td>
<td>7.56</td>
</tr>
</tbody>
</table>
**Rapid gene evolution analysis**

Identification of positively selected genes can reveal insights into unique adaptations. Of the 9,395 1:1 orthologs from the species shown in Fig. 2B, eight genes showed evidence of positive selection in the bowhead whale lineage (Table 7). Similar to comparisons between human and other primates [78], the small number of positively selected genes in the Arctic bowhead whale may reflect a reduced efficacy of natural selection. Indeed, the bowhead whale has a long generation time (~50 years) and a small population size compared to the common minke whale and the other mammals examined here [79]. Unique amino acid changes were recently identified in cetacean haptoglobin (encoded by \( Hp \)) [8], a protein that protects against hemoglobin-driven oxidative stress [80]. \( Hp \) is under positive selection in the bowhead whale (Table 7), suggesting that this gene is rapidly evolving in the long-lived bowhead. Four positively selected genes in the bowhead whale lineage have roles in cancer: mitochondrial tumor suppressor 1 (\( MtuS1 \)) [81]; glycogen synthase kinase 3 alpha (\( Gsk3a \)) and prune exopolyphosphatase (\( Prune \)), which act in concert to regulate cell migration [82]; and cytoplasmic FMR1 interacting protein 1 (\( Cyfip1 \)) [83].

**Concluding comments**

Here, we report the transcriptome of the bowhead whale, the longest-lived mammal known. We compared the gene expression of the bowhead whale to shorter-lived mammals, including the minke whale which we estimate to have diverged from a common ancestor.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Entrez ID</th>
<th>Description</th>
<th>Log2 Fold Change</th>
<th>p-value</th>
<th>adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prkcdbp</td>
<td>ENSMUST0000047040</td>
<td>protein kinase C, delta binding protein</td>
<td>-3.03</td>
<td>1.6E-06</td>
<td>2.3E-04</td>
</tr>
<tr>
<td>Psmg3</td>
<td>ENSMUST0000031531</td>
<td>proteasome (prosome, macropain) assembly chaperone 3</td>
<td>-2.12</td>
<td>3.8E-06</td>
<td>4.3E-04</td>
</tr>
<tr>
<td>Rpa2</td>
<td>ENSMUST00000102561</td>
<td>replication protein A2</td>
<td>-2.61</td>
<td>1.2E-08</td>
<td>1.5E-05</td>
</tr>
<tr>
<td>Rpl10</td>
<td>ENSMUST0000008826</td>
<td>ribosomal protein L10</td>
<td>-2.37</td>
<td>7.7E-07</td>
<td>1.8E-04</td>
</tr>
<tr>
<td>Rps17</td>
<td>ENSMUST0000080813</td>
<td>ribosomal protein S17</td>
<td>-2.34</td>
<td>1.1E-05</td>
<td>7.9E-04</td>
</tr>
<tr>
<td>Scg5</td>
<td>ENSMUST0000024005</td>
<td>secretogranin V</td>
<td>-2.81</td>
<td>6.8E-07</td>
<td>1.7E-04</td>
</tr>
<tr>
<td>Sema4c</td>
<td>ENSMUST00000114991</td>
<td>sema domain, immunoglobulin domain (ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4C</td>
<td>-2.55</td>
<td>1.4E-07</td>
<td>7.7E-05</td>
</tr>
<tr>
<td>Sprtn</td>
<td>ENSMUST0000034467</td>
<td>SprT-like N-terminal domain</td>
<td>-2.14</td>
<td>1.6E-06</td>
<td>2.3E-04</td>
</tr>
<tr>
<td>St7</td>
<td>ENSMUST0000081635</td>
<td>suppression of tumorigenicity 7</td>
<td>2.56</td>
<td>9.5E-06</td>
<td>7.3E-04</td>
</tr>
<tr>
<td>Syf2</td>
<td>ENSMUST0000030622</td>
<td>SYF2 homolog, RNA splicing factor (S. cerevisiae)</td>
<td>2.03</td>
<td>1.2E-05</td>
<td>8.6E-04</td>
</tr>
<tr>
<td>Tle2</td>
<td>ENSMUST00000146358</td>
<td>transducin-like enhancer of split 2, homolog of Drosophila E(sp1)</td>
<td>-2.60</td>
<td>1.6E-06</td>
<td>2.3E-04</td>
</tr>
<tr>
<td>Traf1</td>
<td>ENSMUST00000172159</td>
<td>TNF receptor-associated factor 1</td>
<td>-3.11</td>
<td>1.2E-06</td>
<td>2.2E-04</td>
</tr>
<tr>
<td>Usp9x</td>
<td>ENSMUST0000089302</td>
<td>ubiquitin specific peptidase 9, X chromosome</td>
<td>5.60</td>
<td>5.1E-11</td>
<td>3.7E-07</td>
</tr>
<tr>
<td>Zdhhc1</td>
<td>ENSMUST0000044286</td>
<td>zinc finger, DHHC domain containing 1</td>
<td>-2.16</td>
<td>3.7E-09</td>
<td>8.9E-06</td>
</tr>
</tbody>
</table>
approximately 13-27 million years ago (Mya). This is similar to the great apes (gorillas, chimpanzees, orangutans and humans) which share a common ancestor ~13 Mya [84], or human and rhesus macaques which diverged ~25 Mya [85]. Among higher primates, humans are exceptionally long-lived [86]. Similarly, the bowhead whale has increased maximum lifespan compared to related cetaceans. It has been proposed that the difference in longevity between humans and other primates stems from differential expression of a small number of genes [87]. A recent study comparing humans to eight other mammals, including primates, revealed that 93 liver and 253 kidney genes showed evidence of human lineage-specific expression changes [88]. The number of genes differentially expressed in the bowhead whale liver (45 genes) and kidney (53 genes) compared to other mammals is similar, albeit obtained using a different computational method. We speculate that the genes differentially expressed, with unique coding sequence changes and rapidly evolving in the bowhead whale, represent candidate longevity-promoting genes. We particularly stress the findings suggestive of altered insulin signaling and adaptation to a lipid-rich diet. The availability of a single heart tissue sample from the bowhead whale precluded identification of distinct gene expression patterns in the long-lived bowhead whale, but revealed that argininosuccinate lyase (Asl) may protect the heart of cetaceans during hypoxic events such as diving. The bowhead whale transcriptome provides valuable resource for further research, including genome annotation and analysis, the evolution of longevity, adaptation to an aquatic environment, conservation efforts and as a reference for closely related species.

Table 6. Enrichment of biological process (BP) Gene Ontology (GO) terms for genes differentially expressed in the kidney of bowhead whales compared to other mammals.

<table>
<thead>
<tr>
<th>GO Category</th>
<th>Term</th>
<th>Count</th>
<th>Genes</th>
<th>Fold Enrichment</th>
<th>Fisher's exact P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0000087</td>
<td>M phase of mitotic cell cycle</td>
<td>8</td>
<td>Usp9x, E4f1, Pmf1, Ncapd2</td>
<td>6.1</td>
<td>3.80E-003</td>
</tr>
<tr>
<td>GO:0007067</td>
<td>mitosis</td>
<td>8</td>
<td>Usp9x, E4f1, Pmf1, Ncapd2</td>
<td>6.1</td>
<td>3.80E-003</td>
</tr>
<tr>
<td>GO:000280</td>
<td>nuclear division</td>
<td>8</td>
<td>Usp9x, E4f1, Pmf1, Ncapd2</td>
<td>6.1</td>
<td>3.80E-003</td>
</tr>
<tr>
<td>GO:006350</td>
<td>transcription</td>
<td>20</td>
<td>Egr1, Klf9, Hoxa5, Nr4a1, Tle2, E4f1, Hdac11, Nfya, Pmf1, Junb</td>
<td>2.2</td>
<td>1.10E-002</td>
</tr>
<tr>
<td>GO:0048285</td>
<td>organelle fission</td>
<td>8</td>
<td>Usp9x, E4f1, Pmf1, Ncapd2</td>
<td>5.9</td>
<td>4.40E-003</td>
</tr>
<tr>
<td>GO:000279</td>
<td>M phase</td>
<td>8</td>
<td>Usp9x, E4f1, Pmf1, Ncapd2</td>
<td>4.9</td>
<td>8.40E-003</td>
</tr>
<tr>
<td>GO:000278</td>
<td>mitotic cell cycle</td>
<td>8</td>
<td>Usp9x, E4f1, Pmf1, Ncapd2</td>
<td>4.8</td>
<td>8.80E-003</td>
</tr>
<tr>
<td>GO:0022403</td>
<td>cell cycle phase</td>
<td>8</td>
<td>Usp9x, E4f1, Pmf1, Ncapd2</td>
<td>4.2</td>
<td>1.40E-002</td>
</tr>
<tr>
<td>GO:0051301</td>
<td>cell division</td>
<td>8</td>
<td>Usp9x, E4f1, Pmf1, Ncapd2</td>
<td>4.1</td>
<td>1.60E-002</td>
</tr>
<tr>
<td>GO:0001825</td>
<td>blastocyst formation</td>
<td>4</td>
<td>Nle1, Junb</td>
<td>24.7</td>
<td>2.80E-003</td>
</tr>
<tr>
<td>GO:0045449</td>
<td>regulation of transcription</td>
<td>20</td>
<td>Egr1, Klf9, Hoxa5, Nr4a1, Tle2, E4f1, Hdac11, Nfya, Pmf1, Junb</td>
<td>1.8</td>
<td>4.00E-002</td>
</tr>
</tbody>
</table>
METHODS

Animals. Kidney (n=4), liver (n=3) and heart (n=1) tissues were obtained from bowhead whales (*Balaena mysticetus*) captured during the 2010 native Iñupiaq Eskimo subsistence harvests in Barrow, AK, USA. See Supplemental Table 1 for details. A species range map, which includes all five stocks of bowhead whales, was obtained from the IUCN Red List of Threatened Species (http://www.iucnredlist.org).

Transcriptome sequencing. Total RNA was isolated from frozen tissues using a RNAqueous kit (Life Technologies, Carlsland, CA, USA). Short-insert ‘paired-end’ libraries were prepared using the Illumina TruSeq Sample Preparation Kit v2 and Illumina HiSeq 2000 next-generation sequencing was performed following the manufacturer’s instructions (Illumina, San Diego, CA, USA). Libraries were sequenced bi-directionally (101 bp in each direction) on the Illumina HiSeq 2000 Genome Analyzer platform.

**de novo** transcriptome assembly. We obtained and performed RNA-seq on independent biological replicates of bowhead whales (3 liver, 4 kidney and 1 heart samples) (Supplemental Table 1). In addition, publicly available Illumina HiSeq 2000 RNA-seq data were obtained from the following species: the common minke whale *Balaenoptera acutorostrata* (1 liver, 1 kidney, 1 heart) [8], cow *Bos taurus* (3 livers, 3 kidneys, 3 hearts) [89], (domestic) yak *Bos grunniens* (1 liver, 1 heart) [90], Brandt's bat *Myotis brandtii* (2 livers, 2 kidneys – all from summer active bats) [39], naked mole rat *Heterocephalus glaber* (2 livers, 2 kidneys) [91], Chinese tree shrew *Tupaia chinensis* (1 liver, 1 kidney, 1 heart) [92], mouse *Mus musculus* (3 livers, 3 kidneys, 2 hearts) [93], rat *Rattus norvegicus* (2 livers, 3 kidneys, 2 hearts) [89] and rhesus macaque *Macaca mulatta* (3 livers, 3 kidneys, 2 hearts) [89]. Prior to assembly, all reads (in FASTQ format) were filtered for adapter contamination, ambiguous residues (N’s) and low quality regions using nesoni v0.123 (http://www.vicbioinformatics.com/software.nesoni.shtml) with default settings. Trimmed read quality was visualized using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). For each species, trimmed reads from a tissue were pooled and assembled using Trinity v20140717 [12,13] with default parameters. Assembly statistics can be found in Supplemental Table 2.

Ortholog annotation. To identify orthologous protein-coding transcripts we employed a custom pipeline written in the R statistical computing language [93]. First, we extracted the longest open reading frame for each gene in Mouse Ensembl gene data set GRCm38.p2 (hereafter termed “Reference ORF”) and confirmed they were all in frame, had start and stop codons, and did not have internal stop codons. For genes with multiple alternatively spliced transcripts, the longest transcript was kept. Putative orthologs between the mouse and Trinity assembled transcriptomes (each Trinity transcriptome contains a set of Trinity Transcripts) were identified by a bidirectional best-hit method by discontiguous MegaBLAST. MegaBLAST is bundled in v2.2.29 of the BLAST+ suite and optimized to find long cross-species alignments of highly similar sequences [94]. An “ortholog pair” was declared if a Reference ORF and a Trinity Transcript were mutual best hits and an “ortholog set” was declared if such ortholog pairs could be identified in all species. We further refined the sequences in each ortholog set as follows: they must have a start and stop codons in at least 80% of the sequences in the set; and they must be within ±50% of the median length of the sequences in the set.

**Table 7. Positively selected genes in the bowhead whale.**

<table>
<thead>
<tr>
<th>Gene symbol (mouse)</th>
<th>Ensembl transcript ID</th>
<th>Gene description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC026585</td>
<td>ENSMUST00000046743</td>
<td>quinone oxidoreductase-like protein 2-like</td>
</tr>
<tr>
<td>Cyfip1</td>
<td>ENSMUST00000032629</td>
<td>cytoplasmic FMR1 interacting protein 1</td>
</tr>
<tr>
<td>Ecm1</td>
<td>ENSMUST0000017507</td>
<td>extracellular matrix protein 1</td>
</tr>
<tr>
<td>Gsk3a</td>
<td>ENSMUST00000071739</td>
<td>glycogen synthase kinase 3 alpha</td>
</tr>
<tr>
<td>Gtf2i</td>
<td>ENSMUST00000059042</td>
<td>general transcription factor II I</td>
</tr>
<tr>
<td>Hp</td>
<td>ENSMUST0000074898</td>
<td>haptoglobin</td>
</tr>
<tr>
<td>Mtus1</td>
<td>ENSMUST00000059115</td>
<td>mitochondrial tumor suppressor 1</td>
</tr>
<tr>
<td>Prune</td>
<td>ENSMUST00000015855</td>
<td>prue exopolyphosphatase</td>
</tr>
</tbody>
</table>

Analysis of differential gene expression. For each species, trimmed RNA-seq reads (in FASTQ format) for
each biological replicate were aligned to ortholog sets using TopHat2 [95] and read counting was performed using featureCounts [96]. Raw counts were normalized by Trimmed Mean of M-values (TMM) correction [97] using the R package edgeR [98]. Library size normalized read counts were next subjected to the voom function (variance modeling at the observation-level) function in limma R package v3.18.13 [99] with trend=TRUE for the eBayes function and correction for multiple testing (Benjamini–Hochberg false discovery rate of 0.05). Following limma analysis, strict parameters were set to denote genes as differentially expressed between the bowhead whale and other mammals: significant genes required a log2 fold-change of at least 2.0, a Benjamini Hochberg-adjusted P-value less than or equal to 0.05, and a B-value of at least 2.945; representing a 95% probability that each gene was differentially expressed.

Phylogenetic analysis. Orthologs identified from ten de novo liver transcriptome assemblies were joined into one ‘super gene’ for each species. Briefly, nucleotide sequences were aligned with Clustalo v1.2.1 [100] and trimmed using Gblocks v0.91b, preserving codon information [101, 102]. A ML (Maximal Likelihood) phylogenetic tree was constructed in RAxML-8.0.0 [103] using all codons included and the first and second codon respectively. The best likelihood trees were searched under the GTR-GAMMA model with six categories of rate variation (500 bootstrap replicates were undertaken for estimation of node support). Bayesian molecular dating was adopted to estimate species divergence time using MCMCTree implemented in PAML (v4.4b) [104]. For each Bayesian analysis, 2,000,000 generations of MCMC (Markov chain Monte Carlo) analysis of phylogenetic trees were performed, with the first 2,000 generations discarded as burn-in. The remaining trees were sampled every 100 generations to build consensus trees. Calibration times (50-60 Mya) of divergence between Cetacea (includes the bowhead whale and minke whale) and Artiodactyla (includes the cow and yak) were obtained from fossil records [2, 3].

Identification of proteins with unique amino acid changes. To obtain single best orthologs to human RefSeq proteins, as in the UCSC human 100 species multiple alignment track [105], Trinity FASTA files of the bowhead whale kidney (n=4) and liver (n=3) were concatenated and putative open reading frames (ORFs) identified using the Perl script TransDecoder bundled with Trinity [13]. Next, to associate predicted bowhead whale peptide sequences with human RefSeq IDs, tBLASTn v2.2.29+ of the BLAST+ suite [94] with an E-value cut-off set at 1e-5 was employed. The best match (>50% overall amino acid sequence identity along the entire sequence and spanning >75% of the length of the query sequence) was used to annotate the sequences. Protein sequences from the the toothed whales the bottlenose dolphin (Tursiops truncatus) and the killer whale (Orcinus orca) were available in the UCSC multiple alignment track. The genomes of an additional baleen whale, the minke whale (Balaenoptera acutorostrata) [8] (GenBank Assembly GCF_000493695.1), and two toothed whales: the sperm whale (Physeter catodon) (NCBI Assembly GCF_000472045.1) and the Yangtze River dolphin (Lipotes vexillifer) [11] (GenBank Assembly GCF_000442215.1) were queried using human UCSC multiway coding sequences and gmap v2014-07-28 (a genomic mapping and alignment program for mRNA and EST sequences) [106] with the parameters --cross-species --align --direction=sense force -Y. The gmap output was parsed using custom Perl scripts. Bowhead whale proteins were aligned to orthologs from 67 vertebrate species (Supplemental Table 7) using Clustalo v1.2.1 [100].

In-house Perl scripts were used to parse the Clustalo output and identify unique amino acids. A Perl script scanned orthologous proteins for sites where types/groups of residues unique to the bowhead whale. The script groups residues into four groups: acidic (ED), basic (KHR), cysteine (C) and “other” (STYNQGAVLIFPMW). For example, it identifies cases where the bowhead whale harbors E or D, whilst the other organisms contain exclusively basic, cysteine, or “other” residues at that particular site. The false positive rate of the detection of unique amino acids is approximately 1.33 per analysis; or in other words, one false positive per 315 unique amino acid residue candidates [39]. To validate the results pertaining to specific genes, we manually inspected multiple sequence alignments and interrogated raw bowhead whale transcriptome data for selected genes. We appreciate that, in lieu of lack of bowhead whale genome data, unique amino acid inferences are putative when supported by a limited number of RNA-sequencing reads.

Identification of genes under positive selection in the bowhead whale. The orthologous gene set (9,395 protein-coding genes) identified in our R pipeline was aligned by Clustalo [100]. We next employed Gblocks [101, 102] to minimise the impact of multiple sequence alignment errors and divergent regions. We used the program CodeML in the PAML v4.6 package to perform the optimized branch-site test [104, 107], as described previously [39]. Briefly, we compared PAML modelA1 (where codons evolve neutrally and under
purifying selection) with ModelA (where codons on the branch of interest can be under positive selection). Following PAML analysis, likelihood ratio test (LRT) P-values were computed assuming that the null distribution was a chi-squared distribution with 1 degree of freedom at a false discovery rate of 0.05. We applied a sequential Bonferroni correction to account for the multiple comparisons made in these analyses. Manual inspection of positive selection data is currently recommended in the literature [108, 109], and we manually examined all significant alignments.

ACKNOWLEDGEMENTS

The authors would like to thank the Iñupiaq Eskimo hunters of Barrow for access to captured bowhead whales and for their extraordinary cooperation and involvement. We hope our work is helpful to the conservation of bowhead whales and communities that rely on them. We acknowledge the efforts of Wesley Warren and the Genome Institute at the Washington University School of Medicine, and the Broad Institute in assembling the genome sequence of the sperm whale (Physeter catodon). Next generation sequencing data was submitted to the NCBI Sequence Read Archive (study number PRJNA263931).

Sources of funding

This work was supported by NIH grants AG047745 and AG047200.

Conflict of interest statement

There is no conflict of interest disclosed for any author.

REFERENCES


www.impactaging.com

896

AGING, October 2014, Vol. 6 No.10


SUPPLEMENTAL DATA

Please browse the full text version of this manuscript to see the Supplemental Tables.