Supplementary information

Full Title: The singularity of Cetacea behavior parallels the complete inactivation of melatonin gene modules

Short Title: Melatonin gene modules are eroded in Cetacea

Mónica Lopes-Marques¹•, Raquel Ruivo^{1•*}, Luís Q. Alves^{1,2}, Nelson Sousa¹, André M. Machado¹, L.

Filipe C. Castro^{1,2*}

¹CIIMAR/CIMAR - Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Portugal

²FCUP - Faculty of Sciences, Department of Biology, University of Porto, Portugal

•Equal Contribution

*Correspondence: rruivo@ciimar.up.pt; filipe.castro@ciimar.up.pt

Figure S1: Comparative synteny maps of Aanat genomic locus.	3
Figure S2: Multiple alignment of the predicted exon 1 of Aanat in Cetacea.	4
Figure S3: SRA validation for conserved stop mutation in exon 1 of Aanat in Cetacea.	5
Figure S4: Comparative synteny maps of Asmt genomic locus.	10
Figure S5: Schematic representation of the Human ASMT isoform 1.	11
Figure S6: Multiple alignment of the predicted exon 1 of Asmt in Cetacea.	11
Figure S7: SRA validation for conserved frameshift mutation in exon 1 of Asmt in Cetacea.	12
Figure S8: Alignment of exon 5 of Asmt in multiple mammals.	15
Figure S9: SRA validation for frameshift mutation in exon 4 of <i>P. catodon Asmt</i> .	16
Figure S10: SRA validation for frameshift mutation in exon 3 of <i>L. obliquidens Asmt</i> .	16
Figure S11: Multiple alignment of the predicted <i>Mtnr1a</i> relic sequences (exon 1) in Cetacea.	17
Figure S12: SRA validation for 19 nucleotide deletion in exon 1 of <i>P. catodon Mtnr1a</i> .	17
Figure S 3: Comparative synteny maps of <i>Mtnr1b</i> genomic locus.	18
Figure S14: Multiple alignment of the predicted exon 1 of <i>Mtnr1b</i> in Cetacea.	19
Figure S15: SRA validation for stop mutation in exon 1 of <i>Mntr1b</i> from <i>O. orca</i> .	19
Figure S16: SRA validation for nucleotide insertion generating a premature stop mutation in exon 1 of	20
the Mntr1b gene from B. acutorostrata and E. robustus.	
Figure S17: SRA validation for loss of start codon in exon 1 of the <i>Mntr1b</i> gene from <i>L. obliquidens</i> .	21
Figure S18: Partial multiple alignment of the predicted exon 2 of <i>Mtnr1b</i> in Cetacea.	21
Figure S19: SRA validation for conserved nucleotide deletion in exon 2 of <i>Mtnr1b</i> in Cetacea.	22
Figure S20: SRA validation for nucleotide deletion in exon 2 of <i>Mtnr1b</i> from <i>E. robustus</i> .	25
Figure S21: <i>H. amphibius</i> exon 1 and 5 assembly.	26
Figure S22: Phylogenetic analysis of Perissodactyla, Artiodactyla and Primate Aanat genes.	27
Figure S23: Phylogenetic analysis of Perissodactyla, Artiodactyla and Primate Mtnr1a genes.	28
Figure S24: Phylogenetic analysis of Perissodactyla, Artiodactyla and Primate Mtnr1b genes.	29
Figure S25: T. manatus latirostris gene annotations.	30
Figure S26: SRA validation for loss of start codon in exon 1 of the <i>Aanat</i> gene from <i>T. manatus latirostris</i> .	31
Figure S27: SRA validation for premature stop codon in exon 5 of the <i>Asmt</i> gene from <i>T. manatus latirostris.</i>	31
Figure S28: SRA validation for single nucleotide insertion in exon 2 of the <i>Mtnr1a</i> gene from <i>T. manatus</i>	32
latirostris.	
Figure S29: SRA validation for frameshift mutation in exon 2 of the <i>Mtnr1b</i> gene from <i>T. manatus</i>	32
latirostris.	
Table S1: Blast-n output of Bos Taurus and Homo sapiens sequences against the H. amphibius genome.	33
Table S2: Accession numbers of the Aanat orthologues sequences used in phylogenetic analyses.	34
Table S3: Accession numbers of the <i>Mtnr1a</i> orthologues sequences used in phylogenetic analyses.	35
Table S4: Accession numbers of the <i>Mtnr1b</i> orthologues sequences used in phylogenetic analyses.	36
Gene annotations: In silico gene annotations of Aanat, Mtnr1a and Mtnr1b in H. amphibius	37

AANAT



genome but not annotated

Figure S1: Comparative synteny maps of Aanat genomic locus.



Figure S2: Multiple alignment of the predicted exon 1 of *Aanat* in the listed species. Conserved premature stop mutation, validated with SRA (when available), is represented in the corresponding position with a black arrow.

Orcinus orca

SRA serached SRR1164379- University of Durham 2014-02-13 Sample ID: SAMN02595096 (AR-Genome) SRR574978/81- Baylor College of Medicine 2012-09-20 Sample ID: SAMN01180276 (AForca1)

			1	30			140			16	50					163	
D* BTA Aanat - ex1	ACC	CCA	GAG	ĠAC	GCT	GC	GĠC	GTG	ΤT	rgad	3A I	TG	AG	CGA	5A(ĠĠ	
Frame 2	Т	P	E	D	A	A	G	V	F	E	I		Е	R	E	_	
	2																
Frame 2	т	P	E	D	A	P	G	V	S	E	1		Е	*	E	G	
De REV gnl SRA SRR1164379.97687426.2	ACC	CČA	GAG	GAT	GCT	CG	rggo	GTG	TC	rgao	5A I	TG	AG	TGA	GAC	GGG	FGG
Frame 2	т	P	E	D	A	Р	G	V	S	E	1		E	*	E	G	G
C+ FOD gnl SRA SRR1164379.165109510.1 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	P	lGGC G	CGTG V	TC	rgao E	I AE	TG	AG E	TGA *	GA(E	GGG G	rgg G
Ce REV gnl SRA SRR574981.181257169.2 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CC: P	lGGC G	CGTG V	TC	rgao E	JA I	TG	AG E	TGA	GA(E	GGG S	IGG G
Ce REV gnl SRA SRR574981.194245020.2 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CCS	rggc G	GTG V	TC	rgao E	FA I	TG	AG E	TGA	GA(E	GGG	rgg G
REV gnl SRA SRR574978.124504377.1 Frame 2	ACC T	CCA P	GAG E	GAT D	GCT A	CCS	rggc G	CGTG V	TC	rgao E	FA I	TG	AG E	TGA	GA(E	GGG	IGG
REV gnl SRA SRR574978.147598187.1 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CC 1 P	rggo G	CGTG V	TC	rgao E	GA I	TG	AG E	TGA	GA(E	GGG	IGG
FWD gnl SRA SRR574978.124504377.2 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CCS	rggc G	CGTG V	TC	rgao E	FA I	TG	AG E	TGA	GA(E	GGG	rgg
FWD gnl SRA SRR574978.147598187.2 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CCT	rggo G	CGTG V	TC	rgao E	FA I	TG	AG E	TGA	GA(E	GGG	rgg
FWD gnl SRA SRR1164379.92025766.2 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CCS	rggo G	CGTG V	TCS	rgao E	GA I	TG	AG E	TGA	GA(E	GGG	rgg
Ce FWD gnl SRA SRR574978.52198617.1 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CC: P	rggc G	CGTG V	TCS	rgao E	FA I	TG	AG E	TGA	GA(E	GGG	rgg G
Ce FWD gnl SRA SRR574981.168107312.1 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CC: P	lggc G	CGTG V	TCS	rgao E	FA I	TG	AG E	TGA	GA(E	GGG	IGG G
Ce FWD gnl SRA SRR574981.32364665.1 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CCS	rggc G	CGTG V	TCS	rgao E	FA I	TG	AG E	TGA	GA(E	GGG	IGG G
Ce FWD gnl SRA SRR574978.183270745.1 Frame 2	ACC T	CCA P	.GAG E	G <mark>G</mark> T G	GCT A	CC: P	lege G	CGTG V	TC	r <mark>a</mark> ac K	JA I	TG	AG E	G G G	GA(E	GGG	GG G
Ce REV gnl SRA SRR574978.94024834.2 Frame 2	ACC T	CCA P	.GAG E	GAT D	GCT A	CCT	lGGC G	CGTG V	TC	rgao E	GA I	ΤG	AG E	TGA *	GA(E	GGG S	IGG G
De REV gni SRA SRR574981.110363749.2 Frame 2	ACC T	CCA P	GAG E	GAT D	GCT A	CC1 P	rego G	CGTG V	TC	rgao E	FA I I	TG	AG E	TGA *	GA(E	GGG G	rgg G
Tursiops truncatus																	
Trace Archive Blast																	
ti:1496398209																	
ti:1518196214																	

Ce BTA Aanat - ex1 Frame 2	IN CONTRACTOR DE LA CON	dga R	GAGG
De REV 204523573.scf Frame 2	Tectectracchagestertertectererterter	T GA	GAGGGTGGGTAACCCC
De REV 206776554.scf Frame 2	Tectecetecetesetesetesetesetesetesetesete	TGA TGA	GAGGGTGGGTAACCCC
			'

Tursiops truncatus

SRA searched

SRR606320/21/22/24- Baylor College of Medicine 2015-07-22 Sample ID: SAMN00000070 SRR2148843 - Beijing Genome Institute 2016-08-10 Sample ID: SAMN03968477

		120				130				1	140				150			-			163		
🖙 BTA Aanat - ex1	CT	ĊA	CC	CCA	٩GA	GĠ	AC(GC	Г <mark>G</mark> C	CG:	ĠC	GTO	G TT	ΤG	AGA	Υ	TGA	.60	GΑ	ĠΑ	GĠ		
Frame 1	L		Т	P	E]	D	А	A	۷.	G	V	E		Е	Ι	E		R	E			
	2																						
C+ FWD gnl SRA SRR606320.63088423.1 Frame 1	CT: L	CA	CC T	CCA P	AGA E	GG	ΑT(D	GC! A	rcc P	TG	GC G	GTC V	JAC T	ΤG	AG/ E	Υ	TGA E	GI.	'GA *	GA E	GG	g <mark>G</mark> (GGG G
De REV gnl SRA SRR606324.27615261.1 Frame 1	CTL	CA	CC T	CCA P	AGA E	GG.	ΑT(D	GC! A	rcc P	TG	GC G	GTC V	JAC T	ΤG	AG/ E	Ϋ́	TGA E	GI.	'GA *	GA E	GG	GT(GGG G
C+ FWD gnl SRA SRR606320.78627956.1 Frame 1	CTL	CA	CC T	CCA P	AGA E	GG.	ΑT(D	GC! A	rcc P	TG	GC G	GTC V	JAC T	ΤG	AG/ E	Ϋ́	TGA E	GI.	'GA *	GA E	GG	GT(GGG G
De FWD gnl SRA SRR606324.36465978.1 Frame 1	CTL	CA	CC T	CCA P	AGA E	GG.	A T (D	GC! A	rcc P	TG	GC G	GTC V	JAC T	ΤG	AG/ E	Ϋ́	TGA E	GI.	'GA *	GA E	GG	3 <mark>6</mark> 0 3	GGG G
De REV gnl SRA SRR606324.34872210.1 Frame 1	CT.	CA	CC T	CCA P	AGA E	GG	AT(D	GC! A	rcc P	TG	GC G	GTC V	JAC T	ΤG	AG/ E	Ϋ́	TGA E	GI.	'GA *	GA E	GG(GT(GGG G
De FWD gnl SRA SRR2148843.22763313.1 Frame 1	CT	CA	CC T	CCA P	AGA E	GG/	AT(D	GC! A	rcc P	TG	GC G	GTC V	GAC T	ΤG	AG/ E	ΑT	TGA E	GI.	'GA *	GA E	GG(GT(GGC G
De FWD gnl SRA SRR2148843.155093887.2 Frame 1	CT	CA	CC T	CCA P	AGA E	GG/	AT(D	GC! A	rcc P	TG	GC G	GTO	GAC T	ΤG	AG/ E	A T I	TGA E	G I	'GA *	GA E	GG(3 <mark>G</mark> (GGG G
C+ FWD gnl SRA SRR2148843.148001840.1 Frame 1	CT.	CA	CC T	CCA P	AGA E	GG	AT(D	GC! A	rcc P	TG	GC G	GTO	GAC T	ΤG	AG/ E	Υ	TGA E	GI.	'GA *	GA E	GG(GT(GGG G
C+ FWD gnl SRA SRR2148843.32322684.1 Frame 1	CT	CA	CC T	CCA P	AGA E	GG.	AT(D	GC! A	rcc P	TG	GC G	GTO	GAC T	ΤG	AG/ E	Υ	TGA E	GI.	'GA *	GA E	GG(GT(GGC G
De REV gnl SRA SRR2148843.97235364.2 Frame 1	CT	CA	CC T	CCA P	AGA E	GG.	AT(D	GC! A	rcc P	TG	GC G	GTO	GAC T	ΤG	AG/ E	Υ	TGA E	GI.	'GA *	GA E	GG(GT(GGC G
De REV gnl SRA SRR606320.57758324.1 Frame 1	CT	CA	CC T	CCA P	AGA E	GG.	AT(D	GC! A	rcc P	TG	GC G	GTC V	GAC T	ΤG	AG/ E	Ϋ́	TGA E	GI.	'GA *	GA E	GG	GT(GGG G
De FWD gnl SRA SRR606321.76970487.1 Frame 1	CTL	CA	CC T	CCA P	AGA E	GG	AT(D	GC! A	rcc P	TG	GC G	GTO	GAC T	ΤG	AG/ E	Ϋ́	TGA E	GI.	'GA *	GA E	GG	GT(GGC G
De REV gnl SRA SRR606322.57684769.2 Frame 1	CTL	CA	CC T	CCA P	AGA E	GG.	AT(D	GC! A	rcc P	TG	GC G	GTO	GAC T	ΤG	AG/ E	Ϋ́	TGA E	GI.	'GA *	GA E	GG	GT(GGC G
De FWD gni SRA SRR2148843.36208500.2 Frame 1	CT	CA	CC T	CCA P	AGA E	GG.	AT(D	GC! A	rcc P	TG	GC G	GTO	GAC T	ΤG	AG/ E	Ϋ́	TGA E	GI.	'GA *	GA E	GG	GT(GGC G
C+ FWD gnl SRA SRR2148843.1673738.2 Frame 1	CT	CA	CC T	CCA P	AGA E	GG	AT(D	GC! A	rcc P	TG	GC G	GTO	GAC T	ΤG	AG/ E	A T I	TGA E	GI.	GA *	GA E	GG	GT(GGG G
De REV gnl SRA SRR606323.5237862.2 Frame 1	CT	CA	CC m	CCA	AGA F	GG	AT(GC	rcc p	TG	GC	GTO	GAC T	ΤG	AG/	Υ	TGA F	GI.	GA *	GA F	GG	ET(GGG

Lagenorhynchus obliquidens SRA searched

SRR7349574 - BC Cancer Agency Michael Smith Genome Sciences Centre (BCCAGSC) 2018-06-21 Sample ID: SAMN09386610 SRR7345555 - BC Cancer Agency Michael Smith Genome Sciences Centre (BCCAGSC) 2018-06-21 Sample ID: SAMN09386610

	17		7	1.14				1.00	-7-	1.04			194	100	-
BTA_AANAT_ex1	50 CGTC	GGGATCCC		14	CCAGCGGCG	CCACACAC	cccc		TCCGCTG			GCCGGCGTGTT		CGAGAGG	1
Frame 1 Fro NW_020837988.11:7888117-7891895	P S	G P TGGGATCCC/	G S P	G R AGTCA	CCAGCGGCG	CCACACAC	CCCCAC	CGAGT	F R C	CTCACCCC/	GAGGA GCT	A G V F CCTGGCGTGTC	TGAGATTGAG	GARAGGG	
Frame 1 REV gnl SRA SRR7345555,197719944.2	P P CACC	G I P	AGAGTCCCC	S H AGTCA	Q R R	H T I		C E	F С С Т П СТСС	CTCACCCC	GAGGA GCT	P G V IIS CIGGCGTGTC	E I E TGAGATTGAG	GARAGG	1
Frame 1	PP	G I P		SH	Q R R			E	F C C				E I E		9
Frame 1	P P	G I P	E S P	S H	Q R R	H T I	P T	E	F C C	L T P	E D A	P G V S	E I E	KG	5
Frame 1	PP	G I P	S P	S H	Q R R	H T I	P T	E	F C C	L T P	D A	P G V S	E I E	K G	
Rev gnl SRA SRR7345555.125651147.2 Frame 1	P P	G I P	AGAGTCCCC/	S H	Q R R	H T I	P T	CGAGT	F C C	L T P	GAGGA GCT	C GGCGTGTG C V S	E I E	GARAGGG	4
Fill gnl SRA SRR7345555.426111959.1 Frame 1	P P	G I P		AGTCA S H	Q R R	CCACACAC	P P	CGAGT	F C C	CTCACCCC/	GAGGA GCT	GGCGTGTG G V S	E I E	GALAGG	4
Rev gnl SRA SRR7345555.270380492.1 Frame 1	CACC	G G P	GACTCCCA D S H	AGCCG	CCAGCGGCG Q R R	CCACACAC	P	C GAG	TT GCTGC	CTCACCCC/	GAGGA GCT	CC GGCGTGTC C V S	TGAGATTGAG	GARAGG	
REV gnl SRA SRR7345555.254009893.2 Frame 1		GGGATCCCA G I P		AGTCA	Q R R	CCACACAC		C GAGT	TTTGCTGC	CTCACCCC/	GAGGA GCT	C G G C G T G T G G V S	TGAGATTGAG E I E	GARAGGG	
Feo gnl SRA SRR7345555.394096979.2		GGGATCCC/				CCACACAC	CCCCAC	CGAGT	T GCTGC	CTCACCCC/	GAGGA GCT		TGAGATTGAG	GALAGG	•
Rev gnl SRA SRR7345555.3025957.2		C/				CCACACAC		C GAGT	T GCTGC	CTCACCCC/	GAGGA		TGAGATTGAG	GARAGG	
Rev gnl SRA SRR7345555.424267849.2		C/		SAGTCG	CCAGCGGCG	CCACACAC	CCCCAC	C GAGT	TT GCTGC	CTCACCCC/	GAGGA		TGAGATTGAG	GARAGG	
Rev gnl SRA SRR7345555.3321757.2			TCCCC			CCACAC	CCCCA	C GAGT	TT GCTGC	CTCACCCC	GAGGA		TGAGATTGAG	GARAGG	
REV gnl SRA SRR7345555.78021810.1			Tecce	AGTCA	CCAGCGGCG	CCACACAC	CCCCA	CGAGT	TEGCTGO	CTCACCCC	GAGGA		TGAGATTGAG	GARAGG	-
REV gnl SRA SRR7345555.208139133.1			тессе	AGTCA	CCAGCGGCG	CCACAÇAC	recce	CGAGT	тесто	CTCACCCC/	GAGGA	CEGCCTCT	TGAGATTGAG	GARAGG	2
PEV gnl SRA SRR7345555.393812268.2			cccc		ccaccocco	CCACAÇAC	cccca	CGAGT	тесто	ctcaçcoc	GAGGA	C G G C G T G T G	TGAGATTGAG	GARAGG	-
Prame 1 PEV gnl SRA SRR7345555.448345775.2			ccc		ccăccĝocĝ	CCACAÇAC	cccca	CGAGT	тесто	ctchecce	GAGGA	CT GCC GT GTC	TGAGATTGAG	GARAGG	-
Prame 1 ReV gnl SRA SRR7345555.300019430.2			C/		ccăgcggcg	CCACAÇAC	cccca	C GAGT	тесте	ctcaçccc	GAGGA	CEGCGTGT	TGAGATTGAG	GARAGG	-
Prame 1 P80V gnl SRA SRR7345555.301321443.2				AGTCA	ccăgcggcg	CCACGCAC	reccen	CGAGT	тпосто	ctcagecc	GAGEALGCT	CT GCC T GTC	TGAGATTGAG	GARAGG	2
Frame 1 FEV gnl SRA SRR7345555.275866208.2				GTCA	CCAGCGGCG	CCACAÇAC	CCCCA	CGAGT	тестес	ctcagecc	GAGGA	CEGECGTGT	TGAGATTGAG	GARAGG	-
Frame 1 FeV gnl SRA SRR7345555.390731788.1					CAGCGGCG	CCACAÇAC	rececta	CGAGT	тесте	ctcagecc	GAGGA	CEGCGTGT	TGAGATTGAG	GARAGG	2
Fill gnl SRA SRR7345555.447555052.2					TAGGCG	CCACACAC	rececta c	CGAGT	TEGCTGO	CTCACCCC/	GAGGE		TGAGATTGAG	GALAGG	2
Frame 1 MSV gnl SRA SRR7345555.396550947.2					R R GGCG	CCACACAC		CGAGT	T GCTGC	CTCACCCC/	A A A	G G G S	TGAGATTGAG	GARAGG	-
Frame 1 REV gnl SRA SRR7345555.285159479.2					GGCG	CCACAC		C E	F C C	CTCACCCC	GAGGA GCT	G V S C G G G G G G G C G G C G G V S S V S	E I E TGAGATTGAG	GARAGG	
Frame 1 REV gnl SRA SRR7345555.107506996.2					R	нті	. P 1	E	F C C TGCGCGGG	L T P	GGC GA GCT	G V S GGGCGTGTC	É I E TGAGATTGAG	GARAGG	1
Frame 1 RV en SRA SRR7345555 231519157 1									RG	CTCACCCC	GAGGA GCT	G V S	E I E TGAGATTGAG		1
Parland and and and and and and and and and										CICRECCC!	de l		1 union 11 unio		-

Dr. DTA Assess	122	132	139	149	163
Frame 1	J T P	E D A	TGCCGGC	V F E I E	R E
	>				
Ce FWD NW_020174923.1:7887000-7891000	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	.GTGAGAGGGTGGGCA.
Frame 1		E D A	P G G	V S E I E	* E G G Q
De REV gnl SRA SRR6923830.203433274.2	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	GIGAGAGGGIGGGC
Frame 1		E D A	P G G	V S E I E	* E G G
De FWD gnl SRA SRR6923830.203433274.1 Frame 1	CACCCCA	AGAGGATGC E D A	TCCTGGCGGC P G G	CGTGTCTGAGATTGA V S E I E	GIGAGAGGGIGGGCA
De FIID gni SRA SRR6923830.17914340.1	FCACCCCA	AGAGGATGC	TCCTGGCGGG	CGTGTCTGAGATTGA	.GTGAGAGGGTGGGCA.
Frame 1	J T P	E D A	P G G	V S E I E	* E G G Q
Le FIID gnl SRA SRR6923830.17935068.1	CACCCCA	AGAGGATGC	TCCTGGCGGG	CGTGTCTGAGATTGA	.GTGAGAGGGTGGGCA.
Frame 1		E D A	P G G	V S E I E	* E G G Q
De FOD gnijSRAjSRR6923830.138048032.1 Frame 1	CACCCCA	AGAGGATGC E D A	TCCTGGCGGG P G G	CGTGTCTGAGA <mark>G</mark> TGA V S E <mark>S</mark> E	.GTGAGAGGGTGGGCA
De REV gnl SRA SRR6923830.221582219.1	CACCCCA	AGAGGATGC	TCCTGGCGGG	CGTGTCTGAGATTGA	.GTGAGAGGGTGGGCA.
Frame 1		E D A	P G G	V S E I E	* E G G Q
De REV gnl SRA SRR6923830.87936886.1	CACCCCA	AGAGGATGC	TCCTGGCGGG	CGTGTCTGAGATTGA	GIGAGAGGGIGGGCA.
Frame 1	J T P	E D A	P G G	V S E I E	* E G G Q
De REV gni SRA SRR6923830.122564477.1	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	.GTGAGAGGGTGGGCA.
Frame 1		E D A	P G G	V S E I E	* E G G Q
De REV gnl SRA SRR4292276.90179376.2	CACCCCA	AGAGGATGC	TCCTGGCGGG	CGTGTCTGAGATTGA	.GTGAG
Frame 1	5 T P	E D A	P G G	V S E I E	*
De REV gnl SRA SRR4292276.151941669.2	CACCCCA	AGAGGATGC	TCCTGGCGGG	CGTGTCTGAGATTGA	GIGAGAG
Frame 1		E D A	P G G	V S E I E	* E
De FIID gni SRA SRR6923830.161093567.1	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	GIGAGAGGGIGGGCA.
Frame 1		E D A	P G G	V S E I E	* E G G Q
De FIID gni SRA SRR6923830.221582219.2	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	.GIGAGAGGGIGGGCA.
Frame 1		E D A	P G G	V S E I E	* E G G Q
De FIID gni SRA SRR6923830.122564477.2	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	GIGAGAGGGIGGGCA.
Frame 1		E D A	P G G	V S E I E	* E G G Q
De FUD gni SRA SRR6923830.87936886.2	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	.GIGAGAGGGIGGGCA.
Frame 1		E D A	P G G	V S E I E	* E G G Q
De REV gni SRA SRR4292276.115409536.2	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	CIGAGAGGGIGGGCA
Frame 1	J T P	E D A	P G G	V S E I E	* E G G Q
De FUD gni SRA SRR4292276.136132707.1	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	GIGAGAGGGIGGGCA.
Frame 1	J T P	E D A	P G G	V S E I E	* E G G Q
De FUD gni SRA SRR4292276.138797031.1	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	GIGAGAGGGIGGGCA.
Frame 1	J T P	E D A	P G G	V S E I E	* E G G Q
De FUD gni SRA SRR4292276.141886821.1	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	E TGAGAGGGTGGGCA
Frame 1	J T P	E D A	P G G	V S E I E	
De FWD gnl SRA SRR4292276.61881861.2	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	GIGAGAGGGIGGGCA.
Frame 1	J T P	E D A	P G G	V S E I E	* E G G Q
De REV gnl SRA SRR6923830.338238067.1	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	GIGAGAGGGIGGGCA.
Frame 1	J T P	E D A	P G G	V S E I E	* E G G Q
De FWD gnl SRA SRR4292276.80447047.1	CACCCCA	AGAGGATGC	TCCTGGCGGC	CGTGTCTGAGATTGA	CIGAGAGGGIGGGCA
Frame 1		E D A	P G G	V S E I E	* E G G Q

SRR6923830 - Northwestern Polytechnical University 2018-04-08 Sample ID: SAMN08512128 (MCJT3) SRR4292276- Beijing Genome Institute (BGI) 2017-09-23 Sample ID: SAMN05818630

5197962.127826677.2 GAG<mark>C</mark>AT E H TGAGATIGAGIGAGGGIGGG E I E + E G G Frame 3 P A T M TCS Neophocaena asiaeorientalis SRA searched

Be cancer 18	,		0.110																			
Pt DTA Appet and	100	aa		130		CI m		14	40				1	50		DC A		<i>c</i> .		163		
Frame 3	T	P	AGA	IGGA		A	A		300	V	y T	5	E	GA	I	E		R	E	66		
De REV anIISRAISRR5197962.313848039.2	ACC	CC.	A.G.A	GGI	ΑTG	CTO	CC	ΤGO	GCA	TG	TO	: T (ΞA	GA	ТЛ	GA	ĢΤ	GA	GA	GG	GT	GGG
Frame 3	T	P	E	i i	5	Ă	P	0	3	M	1	3	E	· · ·	Î	E	Υ-	*	Ē	00	G	G
C+ FIID gnl SRA SRR5197962.336869691.2 Frame 3	ACC T	CC. P	A G A E	GGA I I	ΑTG D	CT(A	P	IGO	GCA G	Μ	FT C	CT(S	GA E	GΑ	T] I	GA E	GΤ	GA *	GA E	GG	GT G	GGG G
De FUD gnl SRA SRR5197962.143076514.2 Frame 3	ACC T	CC. P	A G A E	GGA I I	ΑTG D	CT(A	CCS P	TGO	GCA G	M T G	FT C	CT(S	GA E	GΑ	ΤΊ Ι	'GA E	GΤ	GA *	GA E	GG	GT G	GGG G
De FIID gnl SRA SRR5659909.275376799.3 Frame 3	ACC T	CC. P	A G A E	GGA I	ΑTG Ο	CT(A	CC 1 P	TGO	GCA G	M T G	T C	C T (5	GA E	GΑ	ΤΊ Ι	'GA E	GΤ	GA *	GA E	GG	GT G	GGG G
De FUD gnl SRA SRR5659909.275377479.3 Frame 3	ACC T	CC. P	A G A E	GGA I	ΑTG D	CT(A	P	TGO	GCA G	M T G	FT C	C T (GA E	GΑ	T I I	ľGA E	GΤ	GA *	GA E	GG	GT G	G G G G
De FIID gnl SRA SRR5659909.275384071.3 Frame 3	ACC T	CC. P	A G A E	GGA I	ΑTG Ο	CT(A	CC: P	TGO	GCA G	M T G	T C	C T (5	GA E	GΑ	ΤΊ Ι	ľGA E	GΤ	GA *	GA E	GG	G T G	GGG G
De REV gnl SRA SRR5197962.196691459.1 Frame 3	ACC T	CC. P	A G A E	GGA I	λTG D	CT(A	CC: P	TGO	GCA G	M T G	FT C	CT(S	GA E	GΑ	T] I	°GA E	GΤ	GA *	GA E	GG	GT G	GGG
De FIID gnl SRA SRR5659909.88588086.3 Frame 3	ACC T	CC. P	A G A E	GGA I	λTG D	CT(A	CC: P	TGO	GCA G	M T G	T C	CT(S	GA E	GΑ	T] I	'GA E	GΤ	GA *	GA E	GG	G T G	GGG G
De REV gnl SRA SRR5197962.134309602.1 Frame 3	CC P	CC. P	A G A E	GGA I	ΑTG D	CT(A	P	TGO	GCA G	M T G	FT C	CT(S	GA E	GΑ	T I I	'GA E	GΤ	GA *	GA E	GG	GT G	GGG G
De REV gnl SRA SRR5197962.143517536.1 Frame 3	ACC T	C C . P	A G A E	GGA I	λTG Ο	CTO A	CC: P	TGO	GCA G	M T G	T C	C T (GA E	GΑ	T I I	'GA E	GΤ	GA *	GA E	GG	G T G	GGG G
De REV gnl SRA SRR5659909.135106851.3 Frame 3	ACC T	CC. P	A G A E	GGA I	ΑTG D	C <mark>C</mark> (A	CC: P	TGO	GCA G	M T G	G <mark>G</mark> (C T (A	GA E	GΑ	T I I	CGA E	GΤ	GA *	GA E	GG	GT G	GGG G
De FWD gnl SRA SRR5659909.253127781.1 Frame 3	ACC T	CC. P	A G A E	GGA I	λTG Ο	CT(A	P	TGO	GCA G	M T G	T C	C T (GA E	GΑ	T I I	ëGA E	GΤ	GA *	GA E	GG	GT G	GGG G
De FIID gni SRA SRR5659909.435591840.3 Frame 3	ACC T	CC. P	A G A E	GGA I	ΑTG Ο	CT(A	P	TGO	GCA G	M T G	T C	CT(S	GA E	GΑ	T I I	ľGA E	GΤ	GA *	GA E	GG	G <mark>C</mark> G	GGG G
De REV gnl SRA SRR5659909.181457647.1 Frame 3	ACC T	CC. P	A G A E	GGA I	ΑTG	CT(A	P	TGO	GCA G	M T G	FT C	C T (5	GA E	GΑ	T I I	'GA E	GΤ	GA *	GA E	GG	GT G	GGG G
De REV gnl SRA SRR5197962.322478472.2 Frame 3	ACC T	CC. P	A G A E	GGA I	ΑTG D	CT(A	CC: P	TGO	GCA G	M T G	T C	CT(S	GA E	GΑ	T I I	'GA E	GΤ	GA *	GA E	GG	G T G	GGG G
Dt. DEL ##UODALODDE407062427026677.2	100	0.0	8 C 8	C		CI III (0.01	m ~ T		mo		- m -	- 8	C 3	mn	0.01 8	- m	C 2	C 2	~~~	C: m	000

SRA searched SRR5197962 - BC Cancer Agency Michael Smith Genome Sciences Centre 2017-06-27 Sample ID: SAMN06217832 (Qila21) SRR5659909 - BC Cancer Agency Michael Smith Genome Sciences Centre 2017-06-27 Sample ID: SAMN06216270 (Aurora29)

Delphinapterus leucas

Balaenoptera acutorostrata

SRA searched

SRR924087- Korea Institute of Ocean Science and Technology 2013-10-31 Sample ID: SAMN02192644 (MinkeWhale-01) SRR1802582 - Seoul National University 2016-02-11 Sample ID: SAMN03339797 (MW30)

		129		139	1	49		163
De BTA Aanat - ex1	CCAGA	ĠGACO	GCTGC	₫ĠĠĊ <mark>Ġ</mark>	TGT T G.	ÁGATTG	A G C G I	GA <mark>G</mark> Ġ
Frame 1	ΡE	D	A A	G	V F	ΕI	ER	E
	>							
Frame 1	PE	D	A A					
De FIID gnl SRA SRR1802582.53558691.1 Frame 1	CCAGA P E	GGATO D	CTGC A A	TGG				
De REV gnl SRA SRR924087.334666950.1 Frame 1	CCAGA P E	GGATO D	GCTGC A A	TGG				
De FIID gnl SRA SRR924087.353264310.2 Frame 1	CCAGA P E	GGATO D	A A	TGGCA G 1	TG M			
De REV gnl SRA SRR4011121.48255219.2 Frame 1	CCAGA P E	GGATO D	A A	TGGCA G 1	TGTCTG. M S	AGAT E		
De REV gnl SRA SRR1802582.226203039.2 Frame 1	CCAGA P E	GGATO D	A A	TGGCA G I	TGTCTG. M S	AGAT E		
De FIID gnl SRA SRR1802582.88577932.2 Frame 1	CCAGA P E	GGATO D	A A	TGGCA G 1	TGTCTG. M S	AGATT E I		
De REV gnl SRA SRR4011121.75048947.2 Frame 1	CCAGA P E	GGATO D	A A	TGGCA G 1	TGTCTG. M S	AGATTG E I	AGTG E	
De FIID gnl SRA SRR924087.323038446.1 Frame 1	CCAGA P E	GGATO D	CTGC A A	TGGCA G 1	TGTCTG. M S	AGATTG E I	AGTG# E *	GAA E
De FIID gnl SRA SRR1802582.46565943.2 Frame 1	CCAGA P E	GGATO D	CTGC A A	TGGCA G 1	TGTCTG. M S	AGATTG E I	AGTG/ E *	GAAG E
De FIID gnl SRA SRR924087.50174158.1 Frame 1	CCAGA P E	GGATO D	A A	TGGCA G I	TGTCTG M S	GATTG	AGTG/ E *	GAAGGTGGGT/ E G G
De FWD gnl SRA SRR4011121.1449737.1 Frame 1	CCAGA P E	GGATO D	A A	TGGCA G 1	TGTCTG. M S	AGATTG E I	AGTG <i>i</i> E *	GAAGGTGGGT E G G
De REV gnl SRA SRR4011121.55696995.2 Frame 1	CCAGA P E	GGATO D	A A	TGGCA G 1	TGTCTG. M S	AGATTG E I	AGTG# E *	GAAGGTGGGT/ E G G
De REV gnl SRA SRR1802582.113362966.2 Frame 1	CCAGA P E	GGATO D	GCTGC A A	TGGCA G 1	TGTCTG. M S	AGATTG E I	AGTG <i>i</i> E *	GAAGGTGGGT/ E G G
De FIID gni SRA SRR4011121.38205839.2 Frame 1	CCAGA P E	GGATO D	CTGC A A	TGGCA G 1	TGTCTG. M S	AGATTG E I	AGTG/ E *	GAAGGTGGGT/ E G G
De REV gnl SRA SRR4011121.83642666.2 Frame 1	CCAGA P E	GGATO D	GCTGC A A	TGGCA G 1	TGTCTG. M S	AGATTG E I	AGTG/ E *	GAAGGTGGGT/ E G G

Balaenoptera bonaerensis SRA searched

SRR4011113- Institute of Marine Research 2016-08-13 Sample ID: SAMN05447715

De BTA Aanat - ex1 Frame 1	C G R	Cl	C	CCT	120 C J	AC(T	CC	CA(P	G A E	130 G G.	A C	IG (CT (GCI A	GG	140 G C G	G T V	GΊ	F	E G A E	150 G A	Ξ	GAE	G	CGA R	G A E	163 G G		
Fub gnl SRA SRR4011113.193311455.2 Frame 1	CGR	Cl	CGC	CCT L	CP	A C (CC	CA(P	GA E	GG.	A I D	GG	T	GC A	ΤG	G C G	GT	GI	S	E G A	GA	ITI	GA	G	TGA *	G			
Fut gnl SRA SRR4011113.84208481.2 Frame 1	CG R	Cl	CGC	CCT L	CA	A C (T	CC	CA(P	GA E	GG.	A I D	GG	T (GC A	ΤG	G C G	GT V	GΊ	S	E G A	GA	II	GAE	G	TGA *	GΑ			
Fut gnl SRA SRR4011113.57891739.2 Frame 1	CG R	Cl	CGC	CTL	CA	A C (T	CC	CA(P	GA E	GG	A I D	GG	T (GC A	ΤG	G C G	G T V	GI	S	r G A E	GA	ITI	GAE	G	TGA *	G A E	AGO	G T G	G G G
REV gnl SRA SRR4011113.148848400.1 Frame 1	CGR	Cl	C G G	CCT	CP	A C (T	CC	CA(P	GA E	GG.	A I D	GG	T (GC A	ΤG	G C G	GT V	GI	S	E G A	GA	ΤI	GAE	G	TGA *	GA E	AG	G T G	G G G
FWD gnl SRA SRR4011113.30661079.1 Frame 1	CG R	C	CGC	CCT L	CP	A C (T	CC	CA(P	GA E	GG.	A I D	GG	T	GC A	ΤG	G C G	GT V	GΊ	S	E G A	GA	ΤI	GA E	G	TGA *	GA E	AG	G T G	G G G
FWD gnl SRA SRR4011113.201681226.1 Frame 1	CG R	Cl	C G G	CCT L	CA	A C (CC	CA(P	GA E	GG	A I D	GG	CT(GC A	ΤG	G C G	G T V	GI	S	r G A E	GA	II	GAE	G	TGA *	GA E	AG	G T G	G G G
REV gnljSRAJSRR4011113.201681226.2 Frame 1				CTL	CA	T	CC	CA(P	GA E	GG.	A I D	GG	T (GC A	ΤG	G C G	G T V	GΊ	S	E G A E	GA	ΤΊ Ι	GA	G	TGA *	G A E	AG(G T G	G G
					_	_			_		_		_		_			_	_					1					

Eschrichtius robustus

SRA searched

SRR5495100- Purdue University 2017-05-02 Sample ID: SAMN06837694 (GFD-02 -female) SRR5495108 - Purdue University 2017-05-02 Sample ID: SAMN06837692 (ER-14-168 -male)

		130	140	150	163
🖙 BTA Aanat - ex1	.CCCCA	∖GAGĠA <mark>⊂</mark> G(CTGCCGĠCGT(GT <mark>T</mark> TGAĠATTGA(G ⊒ GAGA <mark>G</mark> Ġ
Frame 1	T P	EDA	A G V	FEIE	R E
	>				
REV gnl SRA SRR5495100.168404865.2	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	TGAGAAGGTGGG
Frame 1	T P	E D A	A A G V	S E I E	* E G G
REV gnl SRA SRR5495100.167647318.2	.CCCCP	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	JTGAGAAGGTGCG
Frame 1	T P	E D J	A G V	S E I E	* E G A
Fut gnl SRA SRR5495108.214732318.2	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGA	STGAGAAGGAGGG
Frame 1	T P	E D A	A A G V	S E I E	* E G G
REV gnl SRA SRR5495108.87811596.2	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	STGAGAAGGTGGG
Frame 1	T P	E D A	A A G V	S E I E	* E G G
FWD gnl SRA SRR5495108.269771753.2	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	FGAGAAGGTGGG
Frame 1	T P	E D A	A A G V	S E I E	* E G G
FIID gnl SRA SRR5495108.338730073.1	.CCCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	FGAGAAGGTGGG
Frame 1	T P	E D A	A A G V	S E I E	* E G G
REV gnl SRA SRR5495108.159183030.1	.CCCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	FIGAGAAGGIGGG
Frame 1	T P	E D A	A A G V	S E I E	* E G G
REV gnl SRA SRR5495100.86664043.1	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	STGAGAAGGTGGG
Frame 1	T P	E D A	A G V	S E I E	* E G G
Fut gni SRA SRR5495108.47477046.2	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	STGAGAAGGTGGG
Frame 1	T P	E D J	A G V	S E I E	* E G G
Fut gni SRA SRR5495100.467892117.2	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	STGAGAAGGTGGG
Frame 1	T P	E D A	A G V	S E I E	* E G G
FWD gnl SRA SRR5495100.422453784.2	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	STGAGAAGGTAGG
Frame 1	T P	E D A	A A G V	S E I E	* E G R
REV gnl SRA SRR5495108.222873030.1	.CCCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	FGAGAAGGTGGG
Frame 1	T P	E D A	A A G V	S E I E	* E G G
FIID gnl SRA SRR5495100.457540341.1	.CCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	FIGAGAAGGIGGG
Frame 1	T P	E D A	A G V	S E I E	* E G G
FWD gnl SRA SRR5495100.259032993.1	.CCCCCA	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	TGAGAAGGTGGG
Frame 1	T P	E D A	A G V	S E I E	* E G G
REV gnl SRA SRR5495100.17183361.1	.CCCC <i>P</i>	GAGGATGO	CTGCTGGCGT(GTCTGAGATTGAC	STGAGAAGGTGGG
Frame 1	T P	E D A	A G V	S E I E	* E G G
FID gnlJSRAJSRR5495108.412113239.1	.CCCCA	GAGGATG	CIGCIGGCGI	GTCTGAGATTGA	JTGAGAAGGTGGG * E G G

Balaena mysticetus

SRA searched

SRR1685383 - University of Liverpool 2014-12-31 Sample ID: SAMN03225705 (bowhead whale1) SRR1685385 - University of Liverpool 2014-12-31 Sample ID: SAMN03225705 (bowhead whale1)

De BTA Aanat - ex1 Frame 1	IQCCTCA C L	CCCCAG T P	130 AGGAC E D	GCTGC A A	GGCGTC GV	TTTGAĠATTGA F E I E	GCIGA R	GA <mark>G</mark> Ġ
De FIID gnl SRA SRR1685383.435551008.1 Frame 1	IGCCTCA C L	CCCCAG T P	AGGAT E D	GCTGC	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De FIID gnljSRAJSRR1685383.165632533.2 Frame 1	IGCCTCA C L	CCCCAG T P	AGGAT E D	GCTGC A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De REV gnl SRA SRR1685385.61096953.1 Frame 1	C L	CCCCAG T P	AGGAT E D	GCTGC A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De REV gnl SRA SRR1685385.132892307.1 Frame 1	CCTCA C L	CCCCAG T P	AGGAT E D	GCTGC A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA *	GAAGGTGGG E G G
De REV gnl SRA SRR1685385.113399687.2 Frame 1	C L	CCCCAG T P	AGGAT E D	A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De FIID gnl SRA SRR1685383.309847915.1 Frame 1	C L	CCCCAG T P	AGGAT E D	GCTGC	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De REV gnljSRAJSRR1685383.467306688.2 Frame 1	CCTCA C L	CCCCAG T P	AGGAT E D	GCTGCT A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De FIID gnl SRA SRR1685383.528915563.2 Frame 1	C L	CCCCAG T P	AGGAT E D	GCTGC A A	IGGCGIG G V	ACCGAGCTTGA T E L E	GTGA	GAAGG <mark>G</mark> GGG E G G
De FUD gnl SRA SRR1685383.46706778.2 Frame 1	C L	CCCCAG T P	AGGAT E D	GCTGC A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De REV gnljSRAJSRR1685383.590933776.1 Frame 1	IGCCTCA C L	CCCCAG T P	AGGAT E D	GCTGCT A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De FUD gnl SRA SRR1685383.483786548.2 Frame 1	C L	CCCCAG T P	AGGAT E D	GCTGC A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
C+ FID gnl SRA SRR1685383.571736847.1 Frame 1	C L	CCCCAG T P	AGGAT E D	GCTGC A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA *	GAAGGTGGG E G G
C+ FID gnl SRA SRR1685385.205246094.1 Frame 1	IGCCTCA C L	CCCCAG T P	AGGAT E D	GCTGCT A A	IGGCGIG G V	STCTGAGATTGA S E I E	GTGA	GAAGGTGGG E G G
De FID gnl SRA SRR1685385.315372829.2 Frame 1	C L	CCCCAG T P	AGGAT E D	GCTGC	IGGCGIG G V	STCTGAGATTGA SEIE	GTGA	GAAGGTGGG E G G
De REV gnljSRAJSRR1685383.290581432.2 Frame 1	C L	CCCCAG T P	AGGAT E D	GCTGC A A	IGGCGIG G V	STCTGAGATTGA SEIE	GTGA	GAAGGTGGG E G G
De FIID gnl SRA SRR1685383.559696953.1	IGCCICA	CCCCAG	AGGAT	GÇTGÇ	IGGCGIG	TCTGAGATTGA	GTGA	GAAGGTGGG

Figure S3: SRA validation for conserved stop mutation (red box) in exon 1 of the *Aanat* gene from *O*. *orca*, *T*. *truncatus*, *L*. *obliquidens*, *D*. *leucas*, *N*. *asiaeorientalis*, *B*. *acutorostrata*, *B*. *bonaerensis*, *E*. *robustus* and *B*. *mysticetus*.



Figure S4: Comparative synteny maps of Asmt genomic locus.



Figure S5: Schematic representation of the Human ASMT isoform 1 (not to scale). On top, in blue, relative location of conserved residues in proximity to the hydroxyl side of *N*-acetyl serotonin (NAS), in red residues that establish H-bonds to NAS. On the bottom in orange conserved aromatic residues that encircle the SAM binding site, in red residues that establish H-Bonds to SAM.

			10	20	30	40	50	60	./0
			. .						
H.	sapiens NM_004043.2	ATGGGAT-	CTCAGAG	GACCAGGCC	TATCGCCTCC	TTAATGACTACO	CCAACGGC1	TCATGGTGTC	CCAG
в.	taurus M81862.1	т.с <u></u>		.GTGG.	CA.T	.GGA			
о.	orca	.CT.C	.C.T.G.	GGAA.	.GCG.T	.GG.GAT	G <mark>C</mark>	. T G C G.	
T.	truncatus	.ст.с	.C.T.G.	GGAA.	. GCG . T	.GG.GAT	G <mark>C</mark>	G <mark>C</mark> G.	
D.	leucas	.ст.с	.C.T.G.	GGAA.	.GCG.T	.GG.GA.GTT	G.AC.	G <mark>C</mark> A.	
L.	vexillifer	GAC.TG.C	.C.T.G.	GGA.	.GCG	. GGGG . GA . GT	G <mark>C</mark>	G <mark>C</mark> G.	G.
N.	asiaeorientalis	. CA T. T	.C.T.G.	GGAA.	.GCG.T	.GG.GAT	G <mark>C</mark>	G <mark>C</mark> A.	
в.	acutorostrata	т.с		GGG.	CT	.GGAT	GT.C	c	
в.	bonaerensis	т <mark>.с</mark>		GGG.	CT	.GGAT	GT.C	C	
E.	robustus	т <mark>.с</mark>		GGG.	CT	.GGAT	GT.C	C	
			•						

Figure S6: Multiple alignment of the predicted exon 1 of *Asmt* in the listed species. Conserved single nucleotide insertion, validated with SRA (when available), is represented in the corresponding position with a black arrow.

C+ BTA NM_177493.2 EX1 Frame 1	A C S C C C C C C C C C C C C C C C C C
REV NW_017843847.1:c996537-829636	CIGGCCIGGAGGGAAGACGGGIICCCCIGGGGAGGAGAACIGCGGICICCIGGAGGAATAIICCAGCGCCIICGIGGICGCCCAGGI
Frame 1	A G L E G R I G S P G E N C G L L E Y S S A F V V A O V
C+ REV gnl[SRA[SRR5125024.114636256.1	CTGGCCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGACTGC
Frame 1	W P G G K T G S P G E N C
De REV gnl[SRA[SRR5125024.6902127.1 Frame 1	TTGGCCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCT
Ce REV gnl[SRA[SRR5125024.45501041.2	CIGGCCIGGAGGGAAGACGGGIICCCCCIGGGGAGGAGAACIGCGGICICCIGGAGGAATAIICCAGCGC
Frame 1	A G L E G R I G S P G E N C G L L E Y S S
De REV gnl SRA SRR606319.45788714.1	CTGGCCTGGAGGGAAGACGGGTTCC
Frame 1	W P G G K T G S P G E N C G L
De FWD gnl SRA SRR606319.64109226.1 Frame 1	TTGGCCTGGAGGGAAGACGGGTTCCCCCGGGGAGGAGAACTGCGGTCTCCTGGAG
De FIID gnl[SRA[SRR606319.32549221.2	CTEGCCTEGAGGGAAGACCEGGG G TCCCCCTEGGGAGGAGAACTECCEGGAGG
Frame 1	W P G G K T G S P G E E N C G L L E
De REV gnl SRA SRR606321.34858384.2	CTGGCCTGGAGGGAAGACGGGTTCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGG
Frame 1	W P G G K T G S P G E E N C G L L E
De REV gnl SRA SRR5125024.71308741.2 Frame 1	CIGGCCIGGAGGAGAGACGGGIICCCCIGGGGAGGAGAACIGCGGICICCIGGAGGAATAIICCAGCGCCIICGIGGICGCCCAGGI
C+ F#D gnl SRA SRR5125024.56882281.2	CIGGCCIGGAGGGAAGACGGGIICCCCIGGGGAGGAGACIGCGGICICCIGGGGGAAATAIICCAGCGCCIICGIGGICGCCCCCGGG
Frame 1	A G L E G R I G S P G E E N C G L L A E Y S S A F V V A P G
De REV gnl SRA SRR2148845.39234458.2	CTGGCCTGGAGGGAGAGAGGGTTCC
Frame 1	L A W R E T G S P G E E N C G L
Ce REV gnl SRA SRR5125024.61353609.2	CTGGCCTGGAGGGAAGACGGGTTCC
Frame 1	W P G G K T G S P G E E N C G L L E Y S S A F V V A O V
C+ F#D gnl SRA SRR2148845.136891676.2 Frame 1	CTGGCCTGGAGGGAGAGAGGGTTCCCCCTGGGGAGGAGACTGCGGTCTCCTGGA
De REV gnl SRA SRR606324.48861967.2	CTGGCCTGGAGGGAGAGGGGTTCCC
Frame 1	L A W R E T G S P G E E N C G L L E E Y S S A
C+ F#D gnl SRA SRR606322.86648277.2	CIGGCCIGGAGGGAGACGGGIICCCCIGGGGAGGAGAACIGCGGICICCIGGAGGAATAIICCAGCGCCII
Frame 1	A G L E G R I G S P G E E N C G L L E E Y S S A
De REV gnl SRA SRR5125024.115567367.2	CTGGCCTGGAGGGAGACGGGTGCCCCCGGGGAGGAGAACTGCGGGTCTCCTGGAGGAATATTCCAGCGCCTTCGTGGTCGCCCCAGGT
Frame 1	L A W R E T G P P G E E N C G L L E E Y S S A F V V A Q V
De REV gnl SRA SRR5125024.91035709.1	CIGGCCCGGGAGGAAGACGGGIICCCCCGGGGAGGAGAACIGCGGCICCCGGGGAAAAAIIICCAGCGCCIICGIGGICGCCCAGGI
Frame 1	W P G G K T G S P G E E N C G L L E Y S S A F V V A Q V
De FWD gnl SRA SRR2148845.136274553.2	CTGGCCTGGAGGGAAGACGGGTTCCC
Frame 1	A G L E G R T G S P G E E N C G L L E E Y S
De FWD gnl SRA SRR606321.64782648.1	CTGGCCTGGAGGGAGACGGGTTCCC
Frame 1	L A W R E T G S P G E E N C G L L E E Y S S A F V
De REV gnl SRA SRR606320.14035397.2	CTGGCCTGGAGGGAAGACGGGTTCCC
Frame 1	W P G G K T G S P G E E N C G L L E E Y S S A F V
De REV gnl SRA SRR606319.9519913.1	CTGGCCTGGAGGGAAGACGGGTTCCC
Frame 1	W P G G K T G S P G E E N C G L L E E Y S S A F V V
De REV gnl SRA SRR5125024.16925930.1	CTGGCCTGGAGGGAAGACGGGTTCCC
Frame 1	A G L E G R T G S P G E E N C G L L E E Y S S A F V V A Q V
C+ REV gnl[SRA[SRR2148845.153376043.1	CTGGCCTGGAGGGAAGACGGGTTCC
Frame 1	L A W R E T G S P G E N C G L L E Y S S A F
C+ F#D gnl[SRA[SRR606324.82510474.2	CIGGCCIGGAGGGAAGACGGGIICCCCIGGGGAGGAGAACIGCGGICICCIGGAGGAATAIICAGCGCCIIGGIGGICGCCCCCGGGI
Frame 1	A G L E G R I G S P G E N C G L L E Y F S A L V V A P
De FIID gnl[SRA[SRR606320.3030655.1	CTGGCCTGGAGGAGAGGGGTTCCCCCGGGGAGGAGAACTGCGGGTCTCCTGGAGGAATATTCCAGGGCCTTCGTGGTCGCCCAGGT
Frame 1	L A W R E T G S P G E E N C G L L E E Y S S A F V V A O V
De REV gnl SRA SRR606319.4787579.2 Frame 1	CGEGCCTEGAGGGAAGACGGGTTCCCCGGGGAEGAGAACTGCGGCCTTCGTGGTCGCCCAGGT PCLEGGTCCCCGGGGAEGAGAACTGCGGCCTCCTGGAGGAATATTCCAGCGCCTTCGTGGTCGCCCAGGT PCLEGGTCCCCGGGGAEGAGAACTGCGGTCTCCTGGAGGAATATTCCAGCGCCTTCGTGGTCGCCCAGGT PCLEGGTCGCCCGGGAEGAGAACTGCGGTCTCCTGGAGGAATATTCCAGCGCCTTCGTGGTCGCCCAGGT
De FUD gnl[SRA]SRR606320.31081650.1	<u>TIGGCCTGGAGGGAAGACGGGTTCCCCCGGGGAGGAGAACTGCGGTCTCCTGGAGGAATATTCCAGCGCCTTCGTGGTCGCCCAGGT</u>

SRR606319/20/21/22/24- Baylor College of Medicine 2015-07-22 Sample ID: SAMN00000070 SRR2148845 - Beijing Genome Institute 2016-08-10 Sample ID: SAMN03968479

SRA searched SRR5125024 - National Institute of Standards and Technology 2016-12-27 Sample ID:SAMN06114300 SRR606319/20/21/22/24- Baylor College of Medicine 2015-07-22 Sample ID: SAMN00000070

De BTA NM_177493.2 EX1 Frame 1	A DEGENERACIÓN DE LE CARGA CONCARGE CONCARGE CON CARGA A TADOS CON CONCARGE CON CARGA CONCARGE CON CARGA CONCARGA A TADOS CON A CONCARGA CONCARGA CONCARGA A TADOS CON A CONCARGA CONCA
FILD NW_004438427.1:c7858292-7702561 Frame 1	CTGGAGGGAAGACGGGTTCCC P G G K T G S P G E E N C G L L E E Y S S A F V V A O
Ce REV gnl SRA SRR1164379.173272360.2 Frame 1	CTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGGAGCTGCGGT P G G K T G S P G E E N C G
C+ FWD gnl SRA SRR1164379.34997561.1 Frame 1	CIGGAGGGAAGACGGGICICCCGGGGGGGGGGGGGGGGG
Ce Fuo gnl SRA SRR1164379.169772060.1 Frame 1	CIGGAGGGAAGACGGGICICCIGGGGGGGGGGGGGGGGG
De FIID gnl SRA SRR1164379.130166887.1	CTGGAGGGAAGACGGGTTCCCCTGGGGGGGGAGGAACTGCGGTCTCCTGGAGGAATATTCCAGCGCC
Ce FIID gnl SRA SRR1164379.56907859.1 Frame 1	CTGGAGGGAAGACGGGGTCCCCTGGGGGGGGGGGGGGGG
Ce FIID gnl SRA SRR1164379.182343618.1	CTGGAGGGAAGACGGGTTCCCCCGGGGGGGGAGGAGGAGGAGGGGGGGG
De FIID gnl SRA SRR574977.14178689.1	CTGGAGGGAAGACGGGTTCCCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGGAG
De FIID gnl SRA SRR574982.43781177.1 Frame 1	CTGGAGGGAAGACGGGTTCCCCCTGGGGGAGGAGGAGGAACTGCCGGTCTCCTGGGGGAGGAATATTCCAGCGCCTTTGTGGTCGC P G C K T G S P G E E N C G L L E Y S S A F V V
De FIID gnl SRA SRR574977.9455706.2 Erame 1	CTGGAGGGAAGACGGGTTCCCCTGGGGGAGGAGGAGGAGGAGGGGAGGAGGGCGCCTTTGTGGGCGCC P G C K T G S P G C C N C G L L C Y S S A F V V
Ce FIID gnl SRA SRR574982.57821559.1	CTGGAGGGAAGACGGGTTCCCCCGGGGAGGAGGAGGAGGAGGAGGGAG
De FIID gnl SRA SRR574982.40346974.2	CTGGAGGGAAGACGGGTTCCCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA
De FIID gnl SRA SRR574982.39558494.1	CTGGAGGGAAGACGGGTTCCCCCCGGGGAGGAGGAGGAGGAGGAGGGAG
De FIID gnl SRA SRR574982.31247196.1 Frame 1	CTGGAGGGAAGACGGGTTCCCCTGGGGGAGGAGGAGGAACTGCGGTCTCCTGGGGGAGAATATTCCAGCGCCTTTGTGGTCGC P G C K T G S P G E E N C G L L E E Y S S A F V V
De FIID gnl SRA SRR574982.23324439.2 Erame 1	CTGGAGGGAAGACGGGTTCCCCCCGGGGAGGAGGAGGAGGAACTGCCGGCCCTTGGGGCGCC P G G K T G S P G E E N C G L L E E Y S S A F V V
De FIID gnl SRA SRR574977.67172143.2 Frame 1	CTGGAGGGAAGACGGGTTCCCCCCGGGGAGGAGGAGGAGGAGGAGGGCCTTTGTGGGTCGC P G G R T G S P G G R N C G L L G R P Y S S A F V V
Ce FIID gnl SRA SRR574977.55567757.1	CTGGAGGGAAGACGGGTTCCCCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
Ce FIID gnl SRA SRR574977.31366241.1	CTGGAGGGAAGACGGGTTCCCCCTGGGGGAGGAGGAGGAGGAGGGAG
De FIID gnl SRA SRR574977.26800153.1 Frame 1	CTGGAGGGAAGACGGGTTCCC P G G K T G S P G E E N C G L L E E Y S S A F V V
De FIID gnl SRA SRR574982.122716857.1 Frame 1	CTGGAGGGAAGACGGGTTCCCCCTGGGGGAGGAGGAGGAACTGCGGGTCTCCTGGGGGGAGGAATATTCCAGCGCCTTTGTGGTCGC P G G K T G S P G E R N C G L L E E Y S S A F V V
Ce FWD gnl SRA SRR574982.119918119.2 Frame 1	CIGGAGGGAAGACGGGITCCCCCIGGGGGAGGAGAACIGCGGICTCCCIGGAGGAATAITCCAGCGCCTIIGIGGICGC P G K T G S P G E N C G L E E Y S S A F V V
Ce FILD gnl SRA SRR574982.117739335.1 Frame 1	CIGGAGGGAAGACGGGITCCCCCGGGGGGGGGGGGGGACGCCCCCGGGGGGGGGG
De FIID gnl SRA SRR574982.78332641.1	CTGGAGGGAAGACGGGTTCCCCTGGGGGGGGGGGGGGGAGGACTATTCCAGCGCCTTTGTGGTCGC P G G K T G S P G E E N C G L L E Y S S A F V V
De FIID gnl SRA SRR574982.68487631.2 Frame 1	CIGGAGGGAAGACGGGITCCCCCGGGGAGGAGGAACTGCGGGCCTCCCGGAGGAATATTCCAGCGCCTTIGIGGICGC P G K T G S P G E N C G L E E Y S S A F V V

SRA serached SRR1164379- University of Durham 2014-02-13 Sample ID: SAMN02595096 (AR-Genome) SRR574977/82- Baylor College of Medicine 2012-09-20 Sample ID: SAMN01180276 (AForca1)

Orcinus orca

Tursiops truncatus

Tursiops truncatus

Trace Archive Blast ti:1416172502

De BTA NM_177493.2 EX1 Frame 1	ATGECTCCCCRGCRGCGTGAGGGCTACRGCCTCGRAGGAATABGCCARCGCGTCRTGGTCCCCAG H C S - Q B G G E G Y S L L R E Y A N A F H V S Q
REV NW_017843847.1:c996537-829636 Frame 1	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
De REV 194179132.scf Frame 1	ġĸċġċĸġċĸġċĸġċĸġċĸġċċġċġġġġġĸġċĸġċġġġġġġ

Delphinapterus leucas SRA searched SRR5197962 - BC Cancer Agency Michael Smith Genome Sciences Centre 2017-06-27 Sample ID: SAMN06217832 (Qila21) SRR5197961 - BC Cancer Agency Michael Smith Genome Sciences Centre 2017-06-27 Sample ID: SAMN06216270 (Aurora29)

De BTANM_177493.2 EX1 Frame 3		CONGENEGET GAGGECTACNETCTCCTENAGEAATNOGCCAACGCCTTCNTEETCOCCAE G E G Y S L L R E Y A N A F N V S O
De FIIO gnl[SRA[SRR5197961.41608762.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAA
Frame 3	L E G R T G S	P G E E N C G L L E E
De Feo gnil/SRA/SRR5197961.41583861.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGGGGGGGAACTGCGGTCTCCTGGAGGAA
Frame 3	L E G R T G S	P G E E N C G L L E E
De REV gnl[SRA SRR5197961.88761358.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGGGGGGAACTGCCGGTCTCCTGGAGGAATGTTCCAGCACCT
Frame 3	L E G R T G S	P G E E N C G L L E E C S S T
De REV gnl[SRA[SRR5197961.88719848.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGGGGGGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCT
Frame 3	L E G R T G S	P G E E N C G L L E E C S S T
De REV gnl[SRA[SRR5197962.67877169.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGGGGGGAACTGCCGGTCTCCTGGAGGAATGTTCCAGCACCTTC
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F
De REV gnljSRAjSRR5197962.67877168.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCCGGGCCCCTGGAGGAATGTTCCAGCACCTTC
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F
De REV gnl[SRA[SRR5197962.265361016.2	CCTGGAGGGGAAGACGGGTTC	CCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTC
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F
De FIID gnl[SRA[SRR5197961.301585591.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTG
Frame 3	PGGKTC	P G E E N C G L L E E C S S T F V
Ce FID gnl SRA SRR5197961.301581775.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTG
Frame 3	PGGKTC	P G E E N C G L L E E C S S T F V
C+ FID gnl[SRA[SRR5197961.301586863.2	CCTGGAGGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTG
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F V
De REV gnl[SRA[SRR5197962.381430138.1	CCTGGAGGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTC
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F V V
Ce FID gnl[SRA[SRR5197962.209987741.1	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGGACTGCGGGCTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCC
Frame 3	L E G R T G S	P G E E N C G L L E E C S S T F V V T
C+ FID gnl[SRA[SRR5197961.423578347.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGGCCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCC <mark>G</mark> G
Frame 3	PGGGKTC	P G E E N C G L L E E C S S T F V V T R
Ce Filo gnl[SRA[SRR5197961.423569571.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAG
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F V V T Q
C+ FID gnl[SRA[SRR5197961.423574573.2	CCTGGAGGGGAAGACGGGTTC	CCCCTGGGGAGGAG <mark>G</mark> ACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAG
Frame 3	PGGKTGS	P G E E D C G L L E E C S S T F V V T Q
De REV gnl[SRA[SRR5197961.159483292.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGG
Frame 3	PGGGKTC	P G E E N C G L L E E C S S T F V V T Q V G
De Fill gnl[SRA[SRR5197962.128485458.2 Frame 3	CCTGGAGGGAAGACGGGTTC	C <mark>G</mark> CCTGGGGAGC B GGACT B CGGTCTCCTGGAGGAATGTTCCAGCACC <mark>G</mark> TCG <mark>G</mark> GG <mark>CG</mark> CCCAGGTGGGGGG P G E V D P G L L E E C S S T V G G A Q V G
De REV gnl SRA SRR5197961.255386766.1 Frame 3	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG P G E E N C G L L E E C S S T F V V T Q V G
C+ FID gnl SRA SRR5197962.272497258.2	CCTGGAGGGGAAGACGGGTTC	CCCC <mark>G</mark> GGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3	PGGGKTC	P G E E N C G L L E E C S S T F V V T Q V G G
Ce FID gnl[SRA[SRR5197961.412145489.2	CCTGGAGGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3	PGGKTGS	PGEENCGLLEECSSTFVVTQVGG
Ce FID gnl[SRA[SRR5197961.412510551.2	CCTGGAGGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F V V T Q V G G
De FIID gnl[SRA[SRR5197962.199296168.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCG <mark>G</mark> GGTCACCCAGGTGGGGGG
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F mgm V T Q V G G
Ce FeO gnl SRA SRR5197962.389734764.2	CCTGGAGGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3	PGGGKTGS	P G E E N C G L L E E C S S T F V V T Q V G G
Ce Filo gnl[SRA[SRR5197962.389833925.2	CCTGGAGGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGGTCTCCTGGAGGAATGTTCCAGCACCT <mark>G</mark> CGTGGTCACCCAGGTGGGGGG
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T <mark>C</mark> V V T Q V G G
Ce REV gnl[SRA[SRR5197962.145624657.2	CCTGGAGGGAAGACGGGTTC	CCCCTGGGGAGGAGAACTGCGGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3	PGGGKTC	P G E E N C G L L E E C S S T F V V T Q V G G
De REV gnl SRA SRR5197962.327304011.1	CCTGGAGGGAAGACGGGTTC	CCCTGGGGAGGAGAACTGCGGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3	PGGKTGS	P G E E N C G L L E E C S S T F V V T Q V G G

SRR924087-Korea Inst SRR4011112 - Institute	itute of Ocean Science of Marine Research 20	and Technology 2013-10-31 Sample ID: SAMN02192644 (MinkeWhale-01) 16-08-13 Sample ID: SAMN05447714 (AT)
De BTANM_177493.2 EX1 Frame 1	ATG T GOTC M C S-	CCAGCAGCGCTACAGTCTCCTGAAGGAATAGGCCAAGGCGTTCATGGTCTCCCCAG
REV NW_006725998.1:2215263-2778608 Frame 1	CCTGGAAGGAAGATGGGTTC LEGRMGS	CCCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG P G E E G Y R L L K E Y S S A F M V S Q V G S
De REV gnl SRA SRR924087.465756680.2 Frame 1	CCTGGAAGGAAGATGGGTTC P G R K M G S	CCCTGGGGAGGGGGCTAC P G E E G Y
De FID gnijSRAjSRR4011112.144961281.1 Frame 1	CCTGGAAGGAAGATGGGTTC P G R K M G S	CCCTGGGGAGGGGGCTACCGTCTCCT P G E E G Y R L
De REV gnijSRAjSRR4011112.128790583.1 Frame 1	CCTGGAAGGAAGATGGGTTC P G R K M G S	CCCTGGGGAGGAGGGCTACCGTCTCCTGAAGGA P G E E G Y R L L K
De FID gnljSRAjSRR4011112.108896449.2 Frame 1	CCTGGAAGGAAGATGGGTTC	CCCTGGGGAGGGGGCTACCGTCTCCTGAAGGAATATT P G E E G Y R L L K E Y
De Feb gnljSRAjSRR924087.498164126.2 Frame 1	CCTGGAAGGAAGATGGGTTCC	CCCTGGGGAGGGGGCTACCGTCT#CTGAAGGAATATTCCAGT P G E E G Y R L L K E Y S S
De FID gnljSRAjSRR924087.51926380.2 Frame 1	CCTGGAAGGAAGATGGGTTC P G R K M G S	CCCTGGGGAGGGGGCTACCGTCTCGAGGGAATATTCCAGTG P G E E G Y R L L K E Y S S
De FID gnl SRA SRR4011112.191489128.2 Frame 1	CCTGGAAGGAAGATGGGTTC P G R K M G S	CCCTGGGGAGGAGGGCCACCGTCTCCTGAAGGAATATTCCAGTGC P G E E G H R L L K E Y S S
De FID gnljSRAjSRR924087.211280595.2 Frame 1	CCTGGAAGGAAGATGGGTTC	CCCTGGGGAGGGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCAT T G A P G E E G Y R L L K E Y S S A F I
Ce FID gnl SRA SRR924087.171879263.2 Frame 1	CCTGGAAGGAAGATGGGTTC P G R K M G S	CCCTGGGGAGGGGGCTACCGTCCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGG P G E E G Y R L L K E Y S S A F M V S Q
De REV gnljSRAjSRR924087.84702125.1 Frame 1	CTGGAAGGAAGATGGGTTC	CCCTGGGGAGGGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGT P G E E G Y R L L K E Y S S A F M V S Q
De REV gnl SRA SRR924087.149291345.1 Frame 1	AAGATGGGTTC M G S	CCCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG P G E E G Y R L L K E Y S S A F M V S Q V G S
De REV gnijSRAjSRR924087.433409941.1 Frame 1	L	GGGGAGGAGGGCTACCGTCCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG G E E G Y R L L K E Y S S A F M V S Q V G S
De FID gnl SRA SRR924087.198292309.2 Frame 1		GAGGAGGGCTACCGTCCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG E E G Y R L L K E Y S S A F M V S Q V G S
De Fill gnijSRAjSRR924087.206813918.2 Frame 1		AGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG G Y R L L K E Y S S A F M V S Q V G S
De REV gnl SRA SRR4011112.73202599.2 Frame 1		AGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG G Y R L L K E Y S S A F M V S Q V G S
De FID gnijSRAjSRR924087.33416283.2 Frame 1		GTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG L L K E Y S S A F M V S Q V G S
De Fill gnljSRAjSRR924087.209635720.2 Frame 1		GAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG K E Y S S A F M V S Q V G S
De FID gnijSRAjSRR924087.40277658.1 Frame 1		ATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG Y S S A F M V S Q V G S
De REV gnljSRAJSRR924087.289195282.2 Frame 1		ATATTCCAGTGCCTTCATGGTCTCCCAGGTAGGAAG Y S S A F M V S Q V G S

Balaenoptera acutorostrata

De FIDO gniljSRAJSRR4292276.60895757.2 CTGGAGGGAAGACGGG Frame 1 L E G R T G

me 1 Rev gnl|SRA|SRR6923830.263401553.1 RGACAGG - AAGACGGG na 1 R GACAGG - AAGACGGG R G - E T G

SNA Scarched
SRR924087-Korea Institute of Ocean Science and Technology 2013-10-31 Sample ID: SAMN02192644 (MinkeWhale-01
SPR4011112 Institute of Marine Research 2016 09 12 Sample ID: SAMNI05447714(AT)

De REV gnl[SRA[SRR4292276.8141219.1

Shasearched
SRR924087-Korea Institute of Ocean Science and Technology 2013-10-31 Sample ID: SAMN02192644 (MinkeWhale
SPR4011112 Institute of Marine Research 2016 08 12 Sample ID: SAMNICE 447714/AT)

Frame 1		W	R	1	E	Ť	G	Š		P	G	E	E	Ν		2	Ğ	Ĺ	Ĺ	E	E		Y	S	S	Ā		Ê	Ŷ						
De Fillo gnl SRA SRR6923830.180531754.1 Frame 1	3	rgg W	AG R	GGJ	A A G E	ACC T	GGT	TTC S	CC	CTG P	GGG	E E	GAG E	A A (N	CTC	CG	GTC G	TC L	CIG	GA E	GGA E	A T	A T T Y	CC. S	A G C S	CGC A	CT	TCC F	V V						
De REV gnl SRA SRR6923830.154657412.2 Frame 1	01	rgg W	AG R	GGJ	AAG E	ACC T	GGT	TTC	CC	CTG P	GGG G	E	GAG E	AA(N	CTO	CG	GTC G	TC L	CTG	GA	GGA E	AT.	ATI Y	CC. S	A G C S	CGC A	СT	TCC F	V V	GTO	CAC				
De REV gnl[SRA[SRR6923830.154679079.2 Frame 1	01	rgg W	AG R	GG	A A G E	ACO	GGT	TTC	CC	CTG P	GGG	E A G	GAG	A A (N	CTO	CG	G T C G	TCL	CTG	GAE	GGA E	AT	A T T Y	CC. S	A G C S	CGC A	СT	TC(F	V V	GTO	CAC				
De REV gnl[SRA]SRR6923830.154679664.2 Frame 1	01	rgg W	A G R	GG	A A G E	ACC T	GGT	TTC	CC	CTG P	GGG	AG E	GAG E	AA(N	CTO	GCG	GTC G	TCL	CTG	GA E	GGA E	A T	A T I Y	CC. S	A G C S	CGC A	СT	TC(F	V TG	GTO	CAC				
De REV gnl SRA SRR6923830.8092014.2 Frame 1	31	rgg W	AG	GGJ	A A G E	ACO	GGT	TTC	CC	CTG P	GGG	E	GAG E	A A (CTO	GCG	GTC G	TCL	CTG	GA	GGA E	AT	A T T Y	CC. S	A G C S	CGC A	СT	TCC F	TG V	GTO	CAC				
De REV gnl[SRA]SRR6923830.182577043.2 Frame 1	P	rgg G	AG	GG) G	A A G K	ACC T	GGT	TTC S	CC	CTG P	GGG	AG E	GAG E	A A (N	CTO	GCG	GTC G	TC L	CTGL	GA E	GGA E	AT.	A T T Y	CC. S	A G C S	CGC A	СT	TCC F	Υ V	GTO	CAC T	CCI	A G G Q	,TGO V	30
De REV gnl[SRA[SRR6923830.338048237.1 Frame 1	01	rgg	AG	GGJ G	AAG R	ACC T	GGT	TTC	CC	CTG P	GGG	A G E	GAG	AA N	CTO	GCG	GTC G	TCL	CTG	GAE	GGA E	AT.	A T I Y	CC. S	A G C S	CGC A	СT	TCC F	V TG	GTO	CAC T	CCI	AGG Q	V V	30
De Fillo gnl SRA SRR4292276.158711716.2 Frame 1	C 1	rgg	A G	GG) G	A A G R	ACC T	GGT	TTC	CC	CTG P	GGG	FAG E	GAG E	A A (N	CTO	GCG	G T C G	TCL	CTG	GAE	G G A E	AT	A T T Y	CC. S	A G C S	CGC A	СT	TCC F	V V	GTO	CAC	CCI	A G G Q	,TGO V	30
De Fillo gnlijSRAJSRR4292276.175212994.2 Frame 1	01	FGG	AG	GGJ G	AAG R	AC <mark>Z</mark>	GGT	TTC S	CC	CTG P	GGG	AG E	GAG E	A A (CTO	GCG	GTC G	TC L	CTGL	GA E	GGA E	AT.	A T I Y	CC. S	AG R	GC A	СT	TCC F	Υ V	GTO	CA <mark>A</mark>	CCI	A G G Q	,TGO V	÷(
De Fill gnl SRA SRR4292276.60895757.2	01	rgg	AG	GGI	AAG	ACG	GGT	TTC	CC	CTG	GGG	AG	GAG	AA	CTO	GCG	GTC	TC	CTG	GA	GGA	AT.	ATT	CC.	AGO	CGC	СT	TCO	TG	GTO	CAC	CCA	AGG	TGG	3(

GAGGGAAGACAGG

SRR4292276- Beijing Genome Institute (BGI) 2017-09-23 Sample ID: SAMN05818630 ANGEGE FCCCAGGNGGGEGAGGGCTNCNGTCTCCTGNAGGAATACGCCANCGCGTTCNTGGTCCCCAC

SRA searched SRR6923830 - Northwestern Polytechnical University 2018-04-08 Sample ID: SAMN08512128 (MCJT3)

cccaeeevereeveree a new control of the set o

CTTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATATTCCAGCGCCTTCGTGGTCACCCAGGTGG P G E E N C G L L E Y S A F V V T Q V

AGGIGG

Neophocaena asiaeorientalis

D BTA NM_177493.2 EX1

Eschrichtius robustus SRA searched

SRR5495100 - Purdue University 2017-05-02 Sample ID: SAMN06837694 (GFD-02 -female) SRR5495106 - Purdue University 2017-05-02 Sample ID: SAMN06837692 (ER-14-168 -male)

De BTA NM_177493.2 EX1 Frame 3	ATG T GØTCC M C S	CNGGNGGGGGGGCTACNGTCTCCTGAAGGAATACGCCANCGCGTCTCATGGTCTCCCAG Q B G E G Y S L L K E Y A N A F M V S Q
REV NIPP01008627.1 Frame 3	TGGAAGGAAGATGGGTTCCC	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGGCCTTCATGGTCTCCCAGGTAGG P G E E G Y R L L K E Y S S A F M V S Q V G
C+ FID gnljSRAJSRR5495106.95145541.2 Frame 3	TGGAAGGAAGATGGGTTCCC L E G R M G S	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG P G E E G Y R L L K E Y S S A F M V S Q V G
De REV gnl SRA SRR5495100.334419024.1 Frame 3	TGGAAGGAAGATGGGTTCCC L E G R M G S	CCTGGGGAGGAGGGCTACCGTCTCCT P G E E G Y R L
C+ REV gnl SRA SRR5495100.377301520.1 Frame 3	TGGAAGGAAGATGGGTTCCC	CCCTGGGGAGGAGGGCTACCG <mark>G</mark> CTCCTGAAGGAATA P G E E G Y R L L K E
De REV gnl SRA SRR5495100.164265316.2 Frame 3	TGGAAGGAAGATGGGTTCCC L E G R M G S	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCC P G E E G Y R L L K E Y S
C+ FID gnl SRA SRR5495100.338622623.1 Frame 3	TGGAAGGAAGATGGGTTCCC	CCCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCA P G E E G Y R L L K E Y S
De REV gnIJSRAJSRR5495100.121301504.2 Frame 3	TGGAAGGAAGATGGGTTCCC	CCTGGGGAGGAGGGCTACCGTCTGCAGGAATATTCCAGTGCCTTCATGGTCTCCC P G E E G Y R L L K E Y S S A F M V S
C+ REV gnl SRA SRR5495100.142556638.2 Frame 3	TGGAAGGAAGATGGGTTCCC L E G R M G S	CCCTGGGGAGGAGGGCTACCGTCCTGCAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTA P G E E G Y R L L K E Y S S A F M V S Q V
C+ FID gnljSRAjSRR5495100.397702493.1 Frame 3	TGGAAGGAAGATGGGTTCCC	CCTGGGGANGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG P G BMMP E G Y R L L K E Y S S A F M V S Q V G
C+ FIID gnl SRA SRR5495100.117638356.2 Frame 3	TGGAAGGAAGATGGGTTCCC L E G R M G S	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG P G E E G Y R L L K E Y S S A F M V S Q V G
C+ FID gnljSRAjSRR5495100.134445388.1 Frame 3	TGGAAGGAAGATGGGTTCCC GRKMGS	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCT <mark>G</mark> CCAGG C AGG P G E E G Y R L L K E Y S S A F M V <mark>C</mark> Q A G
De REV gnl SRA SRR5495100.33441463.1 Frame 3	GAAGATGGGTTCCC R M G S	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG P G E E G Y R L L K E Y S S A F M V S Q V G
C+ FID gnl SRA SRR5495100.368367402.1 Frame 3	GAAGATGGGTTCCO R M G S	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG P G E E G Y R L L K E Y S S A F M V S Q V G
C+ FIID gnl SRA SRR5495100.296229175.2 Frame 3	AAGATGGGTTCCC	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG P G E E G Y R L L K E Y S S A F M V S Q V G
C+ FID gnl SRA SRR5495100.157276414.2 Frame 3	TGGGTTCC G S	CCTGGGGAGGAGGGCTACCGTCTCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG P G E E G Y R L L K E Y S S A F M V S Q V G
C+ REV gnl SRA SRR5495100.3343878.1 Frame 3		TCCTGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG L K E Y S S A F M V S Q V G
C+ FIID gnl SRA SRR5495100.81501418.1 Frame 3		TGAAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG K E Y S S A F M V S Q V G
C+ REV gnl SRA SRR5495106.22928576.2 Frame 3		AAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG K E Y S S A F M V S Q V G
Ce REV gnl SRA SRR5495100.368367402.2 Frame 3		AAGGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG K E Y S S A F M V S Q V G
De FID gnljSRAjSRR5495100.472375903.1		GGAATATTCCAGTGCCTTCATGGTCTCCCAGGTAGG

Figure S7: SRA validation for conserved frameshift mutation (red box) in exon 1 of the *Asmt* gene from *O. orca, T. truncatus, D. leucas, N. asiaeorientalis, B. acutorostrata* and *E. robustus.*



Figure S8: Alignment of exon 5 of *Asmt* in multiple mammals, red box indicates conserved Gly187 participating in the active site of ASMT.

Physeter catodon frameshift mutation 1 nt insertion in exon 4 SRA searched

SRR5146843- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 Sample ID: SAMN06187413 SRR5146847- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 Sample ID: SAMN06187412 SRR5146865- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 Sample ID: SAMN06187411



Figure S9: SRA validation for frameshift mutation (red box) in exon 4 of P. catodon Asmt.

Lagenorhynchus obliquidens

129

SRA searched SRR7349574 - BC Cancer Agency Michael Smith Genome Sciences Centre (BCCAGSC) 2018-06-21 Sample ID: SAMN0938661 SRR7345555 - BC Cancer Agency Michael Smith Genome Sciences Centre (BCCAGSC) 2018-06-21 Sample ID: SAMN09386610

	12130	17,130	12,140	17,150	12,160	12328		12,180 12,190
BTA_ASMT_ex3		TGAAGTACATG	GGCAGGACCA		TGGGGCCAC		ACGCCG	130 IGAĠ
Frame 3	S M	K Y M	GRT	S Y R C	WGH	A	DAN	
Frame 3	D M	L Y A	A R T	T Y G	W G H	LA	- A III	V R W V R C
NV gnl [SRA]SRR7345555.319708114.2 Frame 3								
TE gnl [SRA]SRR7345555.374422529.2 Frame 3	GACATGC DI M	TGE GTAC GC G	A R T	CTACC GGG	W G H	C.		20.74
- eni ISRA ISRR7345555.168774541.2 Frame 3	GACATGC DI M		A R T	CTACC GGG	CTGGGGCCAC	CTGGCCG	-CGCGG	G
Terrame 3	GACGTGC		A R T	CTACC GGG	CTGGGGCCACC	L A	-CGCGG	
To gnl JSRA JSRR7345555.361211406.2 Frame 3	GACGTGC	TGETGTACGEG	A R T	CTACC GGG	CTGGGGCCACO		-CGCGG	
W gnl JSRA JSRR7345555.289087344.2 Frame 3	GACATGC D M	TGETGTACGEG	A R T	CTACC GGG	CTGGGGGCCACO W G H	L A	-CGCGG	TGTCAGGTGGGTG
Terrine 3	GACGTGC			CTACC GGG	CTGGGGCCACO W G H			TGTCAGGTGGGGTGGGGGTGT
ISRA SRR7345555.99341020.2 Frame 3	GACGTGC			T Y L G	CTGGGGGCCACO W G H	L A	-CGCGG	TGTEAGGTGGGGTGGGGTGT
49 gnl JSRA JSRR7345555.108925623.2 Frame 3	GACGTGC			CTACC GGG	CTGGGGGCCACO	L A	A I	TGTCAGGTGGGGTGGGGTGT
*By gnl JSRA JSRR7345555.400024801.2 Frame 3	GACATGC D M		A R T	CTACC 666	CTGGGGCCAC	E A	- CGCGGG	TGTCAGGTGGGTGGGGTGT
"Ev gnl JSRA JSRR7345555.396887805.2 Frame 3	GACATGC D M		A R T	CTACC GGG	CTGGGGCCACO W G H	L A	-CSCGG	TGTCAGGTGGGGTGGGGGTGT
Tr gnl JSRA JSRR7345555.18533266.2 Frame a	GACGTGC	TGETGTACGCG	A R T	CTACC GGG	CTGGGGGCCACC W G H	L A	-CGCGG	TGTCAGGTGGGTGGGGTGT
-EV gnl [SRA]SRR7345555.269441146.1 Frame 3	GACATGC D M		A R T		CTGGGGCCACO W G H	CTGGCC G	-CGCGG	TGTCAGGTGGGGTGCGGTGT
** onl ISRA (SRR7345555.343889556.2 Frame a	GACGTGC	TGETGTACGCG			CTGGGGGCCACO W G H		- CGCGG	TGTCAGGTGGGGTGGGGGTGT
"R" gnl [SRA [SRR7345555.270638651.2 Frame 3	GACGTGC	GGETGTACGEG	ALL R T	CTACC GGG	CTGGGGGCCACC	L A	-CGCGG	TGTCAGGTGGGTGGGGTGT
-W gnl JSRA JSRR7345555.132453399.2 Frame 3	GACGTGC			CTACC GGG	CTGGGGCCAC		- CGCGG	TGTCAGGTGGGGTGGGGTGT
"B ¹ gnl [SRA [SRR7345555.98608949.1 Frame 3	GACATGC				CTGGGGCCACO W G H	L A	A	TGTCAGGTGGGTGGGGTGT
FBV gnl [SRA SRR7345555.374422529.1 Frame 3	GACATGC	TGETGTACGEG	A R T	CTACC GGG	CTGGGGCCACO W G H			TGTCAGGTGGGTGCGGTGT

Figure S10: SRA validation for frameshift mutation (red box) in exon 3 of L. obliquidens Asmt.



Figure S11: Multiple alignment of the predicted *Mtnr1a* relic sequences (exon 1) in the listed species. Nucleotide insertion of *P. catodon*, validated with SRA, is represented in the corresponding position with a black arrow

Physeter catodon frameshift mutation 19 nt deletion in exon 1

SRA searched

SRR5146843- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 **Sample ID:** SAMN06187413 **SRR5146847**- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 **Sample ID:** SAMN06187412 **SRR5146865**- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 **Sample ID:** SAMN06187411

	63	73	83	83	103	11	1	123	1	33	143	
BTA_XM_002698656.5:46-1146	AGTGTGCTCAA	GTCTCGCAGO	CCAGCGCCCGGC	GCCC	GAGACGGCGC	GCGGCCGCG	GCCC	CGTGGC	GGCCGCC	CCCTC	GCCTCC	ATCCT
Frame 2	S V L L N	V S Q	PAPG	A	G D G A	R P R	P	S W I	AA	TL	A S	IL
	2											
FIID PCAex1	GCGCTGCTCGA	CGTCTCGCAGO	C <mark>AG</mark> GCGCC <mark>GA</mark> GC	GGCC			CCCI	CGAGGCI	GGCCTCCI	AGCTC	GCCCTC	GTCCT
Frame 2	ALL L D	V S Q	Q A P S	G			P	S R I	AS	R L	A L	V L
Co REV gnllSRAISRR5146865.26111988.1	GCGCTGCTCGA	CGTCTCGCAGO	CAGGCGCCGAGC	GGCC			CCCT	CGAGGCI	GGCCMCCL	AGCTC	GCCCTC	GTC
Frame 2	A L L D	V S Q	Q A P S	G			P	S R I	A S	R L	A L	V
Co. REV apliSRAISRR5146843 16860197 1	GCCCCCCCCCA	CGTCTCGCAGO	AGGCGCCGAGC	GGCC			accer	CGAGGC	GGCC CCC	AGGCTC	GCCCTC	GTCCT
Frame 2	A L L D	V S Q	Q A P S	G			P	S R I	A S	R L	A L	V L
Co. REV gpllSRAISRR5146843 31310687 2			GCCCMGAGC	GGCC			accc1	CGAGGC	GGCC CCC	ANGCTC	GCCCTC	GTCCT
Frame 2			A ? S	G			P	S R I	A S	? L	AL	V L
Co. REV appliSRAISRR5146865 22693162 1			GCGCC <mark>GA</mark> GC	GGCC			CCCT	CGAGGC	GGCC CCC	GGCTC	GCCCTC	GTCCT
Frame 2			A P S	G			P	S R I	A S	R L	AL	V L
De REV gplISRAISRR5146847 20458020 1			GCCGAGC	GGCC			CCCT	CGAGGC	seccadou	GGCTC	GCCCTC	GTCCT
Frame 2			P S	G			P	S R I	A S	R L	AL	V L
De REV gpl/SRAISRR5146843 24980367 1							-		GGCCMCC	GGCTC	GCCCTC	GTCCT
Frame 2									A S	RL	AL	Î Î
De FID opliSRAISRR5146865 19750822 1											CCAC	GTCCT
Frame 2											L	V L
De FIID gnljSRAJSRR5146865.19755989.1											CCTC	GTCCT
Frame 2											L	V L
L							-					

Figure S12: SRA validation for 19 nucleotide deletion in exon 1 of P. catodon Mtnr1a.

MTNR1B



Figure S13: Comparative synteny maps of *Mtnr1b* genomic *locus*.



Figure S14: Multiple alignment of the predicted exon 1 of *Mtnr1b* in the listed species. Nucleotide insertion, generating a premature stop codon, in *B. acutorostrata* and *E. robustus* and premature stop codon retrieved in *O. orca* and *S. chinensis*, mutations validated with SRA (when available), are represented in the corresponding position with a black arrow.



Figure S15: SRA validation for stop mutation (red box) in exon 1 of the Mntr1b gene from O. orca.

Balaenoptera acutorostrata Exon1 1bp insertion creates premature stop codon SRA searched

SRR924087-Korea Institute of Ocean Science and Technology 2013-10-31 Sample ID: SAMN02192644 (MinkeWhale-01) SRR4011112-Institute of Marine Research 2016-08-13 Sample ID: SAMN05447714 (AT)

D+ BTA_NM_001206907.2 - EX1 Frame 3		ATGCC M	GGAC	- A A - E	CGG R	L	R	CCAI	LCTG	L	R	e C (GGGC	R R	CGG P	ecc G	GAGA R	GCC	cccc	CTG L	GAC	W	R G	GCGG R
FID NW_006725477.1:c560939-510360 Frame 3	AGGGTCCAC G S T	ATGCO M I	CCGA0 P E	TAA *	CGG R	L I	PTCG , R	CCA/ Q	ACTG L	CTG L	CGA0 R	GCC	GGGC G	G <mark>A</mark> C R	CGA P	GC <mark>G</mark> S	GAGA G	AGC C	GGG	CTG L	GAC D	TINNN ?	NNAG ?R	G <mark>A</mark> GG R
De REV gnl SRA SRR4011112.75416532.1 Frame 3	G S T	ATGCO M I	CCGAC P E	T AA *	CGG R	CTCCS L I	PTCG	CCA/ Q	ACTG L															
De REV gnl SRA SRR924087.389858527.2 Frame 3	AGGGTCCACC	ATGCO M I	CCGA(P E	TAA *	CGG R	L	PTCG	CCAL	ACTG L	CTG	CGAG	GCC	GGGC G	G <mark>A</mark> C R										
De FIID gnl SRA SRR4011112.161265507.2 Frame 3	AGGGTCCACG G S T	ATGCO M I	CCGA(P E	TAA *	CGG R	CTCC1 L 1	TCG	CCAL	ACTG L	CTG	CGA0 R	GCC	seec e	G <mark>A</mark> C R	CGA P	GC <mark>G</mark> S	GAGA G	GCC	GGG G	CTG	GAC D	TGG <mark>G</mark> W	GCAG	G <mark>A</mark> GG R
De FID gnl SRA SRR4011112.168976966.1 Frame 3		ATGCO M H	CCGAC P E	TAA *	CGG R	CTCCS L 1	PTCG	CCA/ Q	ACTG L	CTG	CGA0 R	GCC	GGGC G	G <mark>A</mark> C R	CGA P	GC <mark>G</mark> S	GAGA G	AGC C	GGG	CTG L	GA <mark>T</mark> D	TGG <mark>G</mark> W	GCAG	G <mark>A</mark> GG R
De REV gnl SRA SRR4011112.187048059.1 Frame 3								CCAP Q	ACTG L	CTG	CGAO R	GCC	GGGC G	G <mark>A</mark> C R	CGA P	GC <mark>G</mark> S	GAGA G	AGC C	GGG	CTG L	GAC	TGG <mark>G</mark> W	GCAG	G <mark>A</mark> GG R
Dr FIID BTA_NM_001206907.2 - EX2																		1						

Eschrichtius robustus Exon1 1bp insertion creates premature stop codon SRA searched

SRR5495100- Purdue University 2017-05-02 Sample ID: SAMN06837694 (GFD-02) SRR5495106- Purdue University 2017-05-02 Sample ID: SAMN06837692 (ER-14-168)

	in the second	1 3	1.0	12		22		32		4	1		62
C+ BTA_NM_001206907.2 - EX1 Frame 3		M P	AC - A	R	LCCTTC	R	LACTG	LGCC	G	GGGC	R P	GGGCC	R
		-		-								_	
Frame 3	R G A G G S T	ATGCCCG M P	E *	R I	TCCTTC L L	R C	ACTGO L	L F	GAGGC	Geec	R P	GAGC	GGAO
REV gnljSRAJSRR5495100.334037093.1 Frame 3	R G A G G S T	ATGCCCG M P	AG AA	AC									
F gnl SRA SRR5495100.126869545.1 Frame 3	R G A G G S T	ATGCCCG M P	AG AA	ACGGC	TCCT L								
FIID gnl SRA SRR5495100.162186712.2 Frame 3	CGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	ATGCCCG M P	AG AA	ACGGC	TCCCTC L P	GC R							
REV gnl SRA SRR5495100.83372043.2 Frame 3	R G A G G S T	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	RCCA	AC						
FIID gnl[SRA SRR5495100.289441073.2 Frame 3	R G A G G S T	ATGCCCG M P	AGTA/	ACGGC	TCCTTC L L	RCCA	AC						
Feo gnl SRA SRR5495100.115867176.1 Frame 3	R G A G G S T	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	RCCA	ACTG						
REV gnl SRA SRR5495100.460523624.2 Frame 3	R G A G G S T	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	GCCA R Q	ACTGO	C T					
REV gnl SRA SRR5495100.280583361.2 Frame 3	R G A G G S T	ATGCCCG M P	AG A	ACGGC	TCCTTC L L	GCCA R C	ACTGO	CTG L					
FIID gnl SRA SRR5495100.287838220.2 Frame 3	CGCGGTGCGGGGGGGGGGGCCACG	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	RCCA	ACTGO	TGCO	*				
FIID gnl SRA SRR5495100.260873243.2 Frame 3	CGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	GCCA	ACTGO	CTGCC L F	GAGGC	2			
FIIO gnl SRA SRR5495100.336766390.2 Frame 3	CGCGGTGCGGGGGGGGGCCACG	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	GCCA	ACTGO	CTGCC L F	AGGC	GGGC	GACC R P	GAGC	G
REV gnl SRA SRR5495100.289441073.1 Frame 3	CGCGGTGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	GCCA	ACTGO	CTGCO L F	AGGC	GGGG	GACC	GAGC	G
REV gnl SRA SRR5495100.145770972.2 Frame 3	R G A G G S T	ATGCCCG M P	AG AA	ACGGC'	TCCTTC L L	RCCA	ACTGO	L F	AGGC	GGGG	GACC R P	GAGC	GG
REV gnl SRA SRR5495100.12325479.1 Frame 3	CGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	ATGCCCG M P	AG AA	ACGGC'	TCCTTC L L	RCCA	ACTGO	CTGCC L F	AGGC	Geec	GACC R P	GAGC	GGA G
FIID gnl SRA SRR5495106.17928474.2 Frame 3	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	GCCA R C	ACTGO	CTGCC L F	AGGC	GGGC	GACC R P	GAGC	GGA G
REV gnl SRA SRR5495100.149525928.2 Frame 3	R G A G G S T	ATGCCCG M P	AG AA	ACGGC	TCCTTC	GCCA R C	ACTGO	TGCC L F	AGGC	GGGC	GACC R P	GAGC	GGA G
FIID gnl SRA SRR5495100.274385853.2 Frame 3	R G A G G S T	ATGCCCG M P	AG AA	ACGGC	TCCTTC L L	R	ACTGO	TGCO L F	GAGGC	GGGG	GACC R P	GAGC	GGA G
FIID gnl SRA SRR5495100.407439466.1 Frame 3	CGCGGTGCGGGGGGGGGGCCACG	ATGCCCG M P	AG TN	R	TCCTTC L L	R	ACTGO	TGCC L F	AGGC	GGGG	GACC R P	GAGC	GGA G
REV gnl SRA SRR5495106.27314627.2 Frame 3	CGCGGTGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	ATGCCCG	AG AA	ACGGC	TCCTTC L L	GCCA	ACTGO	TGCO	AGGC	GGGC	GACC	GAGC	GGA
Feo gnl SRA SRR5495106.74541746.1	NGCGGTGCGGGGGGGGGTCCAC	ATGCCCG	AGTA	CGGC	TCCTTC	GCCA	ACTGO	TGC	AGGC	GGGG	GACC	GAGC	GGA

Figure S16: SRA validation for nucleotide insertion generating a premature stop mutation (red box) in exon 1 of the *Mntr1b* gene from *B. acutorostrata* and *E. robustus*.

Lagenorhynchus obliquidens Loss of start codon in exon 1 of Mtnr1B.

SRA searched

SRR7349574 - BC Cancer Agency Michael Smith Genome Sciences Centre (BCCAGSC) 2018-06-21 Sample ID: SAMN09386610 SRR7345555 - BC Cancer Agency Michael Smith Genome Sciences Centre (BCCAGSC) 2018-06-21 Sample ID: SAMN09386610

	63,270	63,280	63,290	63,300	63,310	63,320	63,330	63,340	63,350	63,360	63,370	63,380
DTA MTHOLD and			1	10	20	29	39	89	59	63	75	м
Frame 1			M	P E N G	S F A	N C C	E A G G	R A E	S P R	WTG	A G	G A R P
FID NW_020838004.1 :2681042-2834039 Frame 1	GETGGEGEGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	S F A	CAACTGCTG(N C C	GAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C	TGGACTGG W T G	GCAGG	GGCGCGCGGGCCC1
REV gnl SRA SRR7345555.254388661.2 Frame 1	G W R G	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	CGAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C	TGGACTAG	G	I
Rev gnl SRA SRR7345555.185579045.2 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGGAGAACGGC P E N G	S F A	CAACTGCTG	CGAGGCGGGCGG E A G G		SAGCC <mark>GGGT</mark> C S R V	TGGACTAG	G	I
Rev gnl SRA SRR7345555.449432141.2 Frame 1	G W R G	A G G S	CACGATA	CCCGGAGAACGGC P E N G	TCCTTCGC 5 F A	CAACTGCTGO N C C	GAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GC	I
Rev gnl SRA SRR7345555.388544591.2 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G		SAGCC <mark>GGGT</mark> C	TGGACTGG	GC	I
Frame 1 SRA SRR7345555.162393719.2	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GC	I
Fe0 gnl SRA SRR7345555.9201283.1 Frame 1	GCTGGCGCGG		CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G		SAGCC <mark>GGGT</mark> C S R V	TGGACTGG	A G	IG
Fito gnl SRA SRR7345555.406533551.2 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCG
RV gnl SRA SRR7345555.63671026.1 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	GAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGCCC1
FRO gnl SRA SRR7345555.324755710.1 Frame 1	GCTGGCGCGGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGGCCC1
Fev gnl SRA SRR7345555.279621960.2 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGCCC1
REV gnl SRA SRR7345555.131357549.2 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	GAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGCCC1
RV gnl SRA SRR7345555.131343183.2 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	GAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGGCCC1
Rev gnl SRA SRR7345555.131339865.2 Frame 1	GCTGGCGCGG		CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G		SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	A G	GGCGCGCGGCCC1
Fito gnl SRA SRR7345555.280472001.2 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC 5 F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGCCC1
RV gnl SRA SRR7345555.134828624.1 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	GAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGGCCC1
REV gnl SRA SRR7345555.89579733.1 Frame 1	GCTGGCGCGGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGGCCC1
Fe0 gnl SRA SRR7345555.264387795.2 Frame 1	GCTGGCGCGG		CACGATA	CCCGGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G		SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	A G	GGCGCGCGGCCC1
REV gnl SRA SRR7345555.49840307.2 Frame 1	G W R G	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC 5 F A	CAACTGCTG	GAGGCGGGCGG E A G G		SAGCC <mark>GGGT</mark> C	TGGACTGG W T G	GCAGG	GGCGCGCGGCCC1
REV gnl SRA SRR7345555.87793784.1 Frame 1	GCTGGCGCGG	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G	CCGAGCGGA	GAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGGGCCC1
REV gnl SRA SRR7345555.249443028.1 Frame 1	W R G	A G G S	CACGATA	CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	CAGGCGGGCGG E A G G	CCGAGCGGA	SAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	GGCGCGCGCGCCC1
FeV gnl SRA SRR7345555.189899623.1 Frame 1	GCGCGG R G	A G G S	CACGATA	CCCGGAGAACGGC P E N G	TCCTTCGC 5 F A	CAACTGCTG	GAGGCGGGCGG E A G G		SAGCC <mark>GGGT</mark> C S R V	TGGACTGG	GCAGG	GGCGCGCGGCCC1
FIO gnl SRA SRR7345555.8749207.1 Frame 1			GTCGAT-(CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	CGAGGCGGGCGG E A G G	CCGAGCGGA	GAGCC <mark>GGGT</mark> C S R V	TGGACTGG W T G	GCAGG	
Fito gnl SRA SRR7345555.8748298.1 Frame 1		C S L M	GTCGAT-(CCCGAGAACGGC P E N G	TCCTTCGC S F A	CAACTGCTG	GAGGCGGGCGG E A G G		SAGCC <mark>GGGT</mark> C	TGGACTGG	GCAGG	GGCGCGCGGCCC1
Rev gnl SRA SRR7345555.369211374.2	G	GCGGGAGGAT	CACGATA	CCEGAGAACGGC	TCCTTCGC	CAACTGCTG	GAGGCGGGCGG	CCG R GC <mark>G</mark> GA	SAGCC <mark>GGGT</mark> C	TGGACTGG	GCAGG	legeeçeegeeçet

Figure S17: SRA validation for loss of start codon in exon 1 of the *Mntr1b* gene from *L. obliquidens*.



Figure S18: Partial multiple alignment of the predicted exon 2 of *Mtnr1b* in the listed species. Conserved nucleotide deletion, validated with SRA (when available), is represented in the corresponding position with a black arrow.

Delphinapterus leucas conserved 1 nt deletion exon 2

SRA searched

SRR5197962 - BC Cancer Agency Michael Smith Genome Sciences Centre 2017-06-27 Sample ID: SAMN06217832 (Qila21) SRR5659909 - BC Cancer Agency Michael Smith Genome Sciences Centre 2017-06-27 Sample ID: SAMN06216270 (Aurora29)

C. BTA_NM_001206907.2 - EX2	CGCCGGA	ACACGG	IGGCCGT	GTGCTC	GTGCAC	TTEETG	G	CCCGT	GÓCTGI	TGTGTG	TTTCTG	CTACCT	GCACA	TETGGG	GCTGG
Frame 3	AG	ITP		VL	V A	E 14		PV	A		EC	1 1	A		
C+ FID NW_019160975.1:c1776641-1743906 Frame 3	A R	ACACGG	GGCAGTO A V	GGTG <mark>G</mark> TC	GTCCAC V H	TTCCTC F L	CIC - E -	CCCAC P T	GGCTG1 A V	TGTGTG V C	CTTCTG F C	TATCT	GC <mark>G</mark> CA	CTGGG	rgcTgg / L
REV gnl SRA SRR5197962.92138182.1 Frame 3	A R	ACACGG	GGCAGTO A V	GTG <mark>G</mark> TC V	GTCCAC V H	TTCCTC F L	CIC - E -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C				
FeD gnl[SRA[SRR5197962.164475351.2 Frame 3	AR	ACACGG	GGCAGTO A V	GGTG <mark>G</mark> TC	GTOCAC V H	TTCCTC F L	ିଲ୍ଲ - ଜୁନ	CCCAC P T	GGCTG1 A V	TGTGTG V C	TTCTG F C	12			
REV gnl[SRA[SRR5197962.79593361.1 Frame 3	A R	ACACGG	GGCAGTO A V	GGTGGTC	GTCCAC V H	TTCCTC F L	CIC - E -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TA			
FID gnl SRA SRR5197961.97140020.1 Frame 3	A R	ACACGG	GGCAGTO A V	GGTG <mark>G</mark> TC V V	GTOCAC V H	TTCCTC F L	CIC - F -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TA			
For gnl[SRA[SRR5197962.266541145.2 Frame 3	A R	ACACGG	GGCAGTO A V	GGTGGTC	GTOCAC V H	TTCCTC F L	이다 - 프 -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TA			
REV gnl SRA SRR5197961.51796410.2 Frame 3	A R	ACACGG	GGCAGTO A V	GGTGGTC	GTCCAC V H	TTCCTC F L	이면 - - -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TACCT	GC		
Rev gnl SRA SRR5197961.207976215.1 Frame 3	A R	ACACGG	GGCAGTO A V	GTGGTC	GTOCAC V H	TTCCTC F L	CIC - F -	CCCAC P T	GGCTGI A V	TGTGTG	TTCTG F C	TACCT	GC		
FID gnl SRA SRR5197962.219523073.1 Frame 3	A R	ACACGG	GGCAGTO A V	GTCCTC	GTOCAC V H	TTCCTC F L	이다. - 프 -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TACCT Y L	GC <mark>G</mark> C R		
FID gnl SRA SRR5197961.77436595.2 Frame 3	A R	ACACGG	GGCAGTO A V	GGTG <mark>G</mark> TC	GTCCAC V H	TTCCTC F L	CIC - E -	CCCAC P T	GGCTGT A V	TGTGTG V C	TTCTG F C	TATCT	GCGCA	CTGG T W	
F#D gnl SRA SRR5197962.355214741.1 Frame 3	A R	ACACGG	GGCAGTO A V	V V	GTOCAC V H	TTCCTC F L	이다 - 프 -	CCCAC P T	GGCTGT A V	TGTGTG V C	TTCTG F C	TACCT	GCGCA	T TGGG	rG /
REV gnljSRAJSRR5197961.34392783.1 Frame 3	A R	ACACGG	GGCAGTO A V	GGTGGTC	GTOCAC V H	TTCCTC F L	CIC -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TACCT Y L	GCGCA	CTGGG'	rgc /
REV gnl SRA SRR5197961.211682075.2 Frame 3	A R	ACACGG	GGCAGTO A V	V V	GTCCAC V H	TTCCTC F L	CIC - E -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TTCCT F L	GCGCA R	CTGGG	FGCTG / L
F#D gnljSRAjSRR5197962.81011082.1 Frame 3	A R	ACACGG	A V	V V	GTOCAC V H	TTCCTC F L	C C -	CCCAC P T	GGCTGI A V	TGTGTG	TTCTG F C	TACCT	GCGCA		FGCTGG
REV gnl SRA SRR5197961.155956791.1 Frame 3	A R	ACACGG	GGCAGTO A V	V V	GTCCAC V H	TTCCTC F L	CIC - E -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TACCT Y L	GC <mark>G</mark> CA	T W	FGCTGG
REV gnl SRA SRR5197961.155911959.1 Frame 3	A R	ACACGG	GGCAGTO A V	V V	GTCCAC V H	TTCCTC F L	C C - E -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TACCT Y L	GC <mark>G</mark> CA	CTGGG	FGCTGG / L
REV gnl SRA SRR5197961.156060595.1 Frame 3	AR	ACACGG	GGCAGTO A V	V V	GTCCAC V H	TTCCTC F L	CIC - E -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG	TACCT	GC <mark>G</mark> CA	CTGGG	FGCTGG
FeD gnl SRA SRR5197962.61762039.1 Frame 3	A R	ACACGG	GGCAGTO A V	V V	GTCCAC V H	TTCCTC F L	ି <u>ଜ</u> - ଜୁ -	CCCAC P T	GGCTG1	TGTGTG V C	TTCTG F C	TATCT	GC <mark>G</mark> CA	T C	FGCTGG
REV gnlJSRAJSRR5197961.128279279.2 Frame 3	A R	ACACGG	GGCAGTO A V	GGTG <mark>G</mark> TC	GTCCAC V H	TTCCTC F L	C C - E -	CCCAC P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TATCT	GC <mark>G</mark> CA R	T W	EGCTGG / L
FID gnl[SRA[SRR5197961.322378481.1 Frame 3	A R	ACACGG	GGCAGTO A V	GGTG <mark>G</mark> TC V V	GTCCAC V H	TTCCTC F L	C C - E -	P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TATCT	GC <mark>G</mark> CA	CTGGG	rgcTgg / L
Fuo gnljSRAJSRR5197961.346090379.1 Frame 3	AR	ACACGG	GGCAGTO A V	V V	GTOCAC V H	TTCCTC F L	CIC - E -	P T	GGCTGI A V	TGTGTG V C	TTCTG F C	TATCT	GCGCA R	CTGGG	FGCTGG
F#D gnl SRA SRR5197961.276824122.2 Frame 3	A R	ACACGG	GGCAGTO A V	GGTG <mark>G</mark> TC	GTCCAC V H	FL	이다. - 프 -	CCCAC P T	GGCTG1 A V	TGTGTG V W	GTTCTG F C	TATCT	GC <mark>G</mark> CA	CIGGG	EGCTGG

Neophocaena asiaeorientalis conserved 1 nt deletion exon 2

SRA searched

SRR6923630- Northwestern Polytechnical University 2018-04-08 Sample ID: SAMN08512128 SRR4292276- Beijing Genome Institute 2017-09-23 Sample ID: SAMN05818630

	593	603		613	623		A33		643	653		662	672	
De BTA_NM_001206907.2 - EX2 Frame 3	A G	Y T	A A V	GTGCTCG	H	CCTGC	GCC	C <mark>GTGGC</mark> V A		C F	C - Y	L H	ATCTGGGT	
De Fillo NW_020175536.1:c3825900-3812000 Frame 3	A R	TACACGG Y T	TGGCAGTG A V	GT <mark>TG</mark> TCG: VVV	TCCACT: V H I	FCCTCC	- C C (CACGGC:	V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	ACCTGGGT T W V	-GCTG - L
De REV gnljSRAJSRR6923830.127049872.1 Frame 3	A R	TACACGG Y T	GGCMGTG A V	GT <mark>TG</mark> TCG	TCCACT V H I	FCCTCC	- CC) - P	CACGGC:	P					
De REV gnl SRA SRR6923830.265314935.1 Frame 3	A R	TACACGG Y T	GGCMGTG A V	GT <mark>TG</mark> TCG	V H I	FCCTCC	- CC) - P	CACGGC!	IGTTG V					
De REV gnl SRA SRR6923830.268333416.2 Frame 3	A R	TACACGG Y T	GGCMGTG A V	GT <mark>TG</mark> TCG V V V	V H I	FCCTCC	- C C (CACGGC!	IGTTG V					
De Fillo gnijSRAjSRR6923830.331146485.1 Frame 3	A R	TACACGG Y T	GGCAGTG	GT <mark>TG</mark> TCG	TCACT V H I	FCCTCC	-CC) - P	CACGGC1	IGTIGIG V V	,				
C+ F=0 gnl SRA SRR6923830.127616286.1 Frame 3	A R	TACACGG Y T	GGCAGIG A V	GTTGTCG	V H I	FCCTCC	- C C (CACGGC ?	IGTIGIG V V	TGCTTC C F	TG <mark>T</mark> -TACO C - Y	L R	T W V	-GCTG - L
C+ F=0 gnl SRA SRR6923830.57756137.2 Frame 3	A R	TACACGG Y T	GGCAGIG A V	GTTGTCG	V H I	FCCTCC	- C C (CACGGC!	IGTIGIG V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	L R	T W V	- GCTG - L
C+ F=0 gnl SRA SRR6923830.189545986.2 Frame 3	A R	TACACGG Y T	GGCAGIG A V	GTTGTCG	V H I	FCCTCC	- C C (CACGGC!	IGTIGIG V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	L R	T W V	-GCTG - L
C+ REV gnl SRA SRR6923830.308628213.2 Frame 3	A R	TACACGG Y T	GGCAGTG	GTTGTCG	V H I	FCCTCC	- CC) - P	CACGAC:	IGTIGIG V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	T W V	-GCCT - P
De REV gnl SRA SRR6923830.117776647.1 Frame 3	A R	TACACGG Y T	GGCMGTG A V	GTTGTCG	TCACT: V H I	FCCTCC	- C C (CACGGC ! T A	IGTIGIG V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	ACCIGGGI T W V	-GCTG - L
De REV gnl SRA SRR4292276.13508772.1 Frame 3	A R	TACACGG Y T	GGCAGIG A V	GTTGTCG	V H I	FCCTCC	- C C (CACGGC!	r					
De Fillo gnijSRAjSRR6923830.117776647.2 Frame 3	A R	TACACGG Y T	GGCAGIG A V	GTTGTCGT VVV	TCACT V H I	FCCTCC	- CC) - P	CACGGC!	IGTIGIG V V	C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	ACCTGGGT T W V	-GCTG - L
De Fillo gnijSRAjSRR6923830.152122811.1 Frame 3	AGACAGO D R	TACACGG Y T	GGCAGIG A V	GTTGTCG	V H I	FCCTCC	- CC) - P	CACGGC :	IGTIGIG V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	ACCTGGGT T W V	-GCTG - L
De REV gnl SRA SRR6923830.100104035.2 Frame 3	A R	TACACGG Y T	GGCAGTG A V	GTTGTCG	V H I	FCCTCC FL	- CC) - P	CACGGC ?	IGTIGIG V V	R F	TGT-TACO C - Y	L R	T W V	-GCTG - L
De Fillo gnijSRAjSRR4292276.60672204.2 Frame 3	A R	TACACGG Y T	GGCAGTG	GTTGTCG	V H I	FCCTCC FL	- CC) - P	CACGGC	V V	C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	T W V	-GCTG - L
De REV gnl SRA SRR6923830.127616286.2 Frame 3	CCCCG R	TACACGG Y T	GGCAGIG	GTTGTCG	V H I	FCCTCC FL	- CC) - P	CACGGC:	V V	C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	ACCTGGGT T W V	-GCTG - L
De REV gnl SRA SRR6923830.152122811.2 Frame 3	GG	TACACGG Y T	GGCAGTG	GTTGTCG	V H I	FCCTCC	- CC) - P	CACGGC:	V V	C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	T W V	-GCTG - L
De Fillo gniljSRAJSRR4292276.20056905.2 Frame 3		ACACGG T	GGCAGTG	GTTGTCG	V H I	FCCTCC	- CC) - P	CACGGC:	V V	C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	T W V	-GCTG - L
De REV gnl SRA SRR4292276.40202837.2 Frame 3		CACGG T	IGGC <mark>A</mark> GIG A V	GTTGTCG	V H I	FCCTCC	- CC) - P	CACGGC	V V	C F	TG <mark>T</mark> -TACO C - Y	L R	T W V	-GCTG - L
C+ REV gnl SRA SRR6923830.244111324.1 Frame 3			GGC <mark>M</mark> GTG A V	GTTGTCG	V H I	FCCTCC	- CCI - P	T A	V V	C F	TG <mark>T</mark> -TACO C - Y	L R	T W V	-GCTG - L
C+ F=0 gnl SRA SRR6923830.304046029.1 Frame 3			GC <mark>A</mark> GTG A V	GTTGTCG	V H I	FCCTCC	- CC) - P	CACGGC ?	V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	CTGC <mark>G</mark> C L R	T W V	-GCTG - L
Ce REV gnl SRA SRR6923830.285837699.1 Frame 3			GG	GTTGTCG	V H I	FCCTCC	- CC) - P	CACGGC ?	V V	GTGCTTC C F	TGT-TAC C - Y	TGC <mark>G</mark> C V R	T W V	-GCTG - L
Ce REV gnl SRA SRR6923830.140433164.2 Frame 3			GG	GTTGTCG	TCCACT V H	CCTCC L	- CC) - P	CACGGC?	IGTIGIG V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	L R	T W V	-GCTG L
De REV gnl SRA SRR6923830.273703277.1 Frame 3			GG	GTTGTCG	TCCACT V H I	FCCTCC FL	- CC) - P	CACGGC!	V V	GTGCTTC C F	TG <mark>T</mark> -TACO C - Y	L R	T W V	-GCTG - L

Physeter catodon 1 nt deletion and stop codon exon 2

SRA searched

SRR5146843- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 Sample ID: SAMN06187413 SRR5146865- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 Sample ID: SAMN06187411

	385	395	405	415		424	434	444	454
DFA_NM_001206907.2 - EX2	TTCAACA	CACCGCC	ATCGCCGTCG	ACCGCTACTGCT	-	ACGTCTG	CCGCAGCGTGAC	TAC	CACCGCCTCTGC
Frame 3	F N	ITA	IAV	DRYC	-	Y V C	R S V T	Y	HRLC
	2								(
De FUD NW_019873589.1:20367954-20391	TTC <mark>G</mark> ACA	TCACCGCC	ATCGCCATTA	ACCGCTACTGCI	T	ACGTCTG	CCACAGCGTGAC	TAA	CACCGCATCTAC
Frame 3	F D	ΙΤΑ	IAI	N R Y C	L	Y V C	H S V T	*	HRIY
FIID gnl SRA SRR5146865.19157539.1	TTC <mark>G</mark> ACA	TCACCGCC	ATCGCCATTA	ACCGCTACTGCI	T	ACGTCTG	CCACAGCGTGAC	TAA	CACCGCATCTAC
Frame 3	F D	ΙΤΑ	IAI	NRYC	L	Y V C	H S V T	*	HRIY
REV gnl SRA SRR5146865.915291.1	TTCGACA	TCACCGCC	ATCGCCATAA	ACCGCTACTGCI	т	ACGTCTG	CCACAGCGTGAC	TAA	CACCGCATCTAC
Frame 3	F D	ΙΤΑ	IAI	NRYC	L	Y V C	H S V T	*	HR IY
REV gnl[SRA]SRR5146843.26606333.1			ATCGCCATTA	ACCGCTACTACI	т	ACGTCTG	CCACAGCGTGAC	TAA	CACCGCATCTAC
Frame 3			IAI	N R Y Y	L	Y V C	H S V T	*	H R I Y
FIID gnl SRA SRR5146843.33070640.2			CCATTA	ACCGCTACTGCI	т	ACGMCTG	CCACAGCGTGAC	TAA	CACCGCATCTAC
Frame 3			I	NRYC	L	Y ? C	H S V T	*	H R I Y
FIID gnl SRA SRR5146843.33083365.2			CCATTA	ACCGCTACTGCI	т	ACGMCTG	CCACAGCGTGAC	TAA	CACCGCATCTAC
Frame 3			I	NRYC	L	Y ? C	H S V T	*	HRIY
FIID gnl SRA SRR5146843.16333760.1						TCTG	CCACCGTGAC	TAA	CACCGCATCTAC
Frame 3						C	H S V T	*	H R I Y
Fut gnllSRAJSRR5146865.12398546.1						TCTG	CCACAGCGTGAC	TAA	CACCGCATCTAC
Frame 3					Ш	C	H S V T	*	HRIY
REV gnllSRAJSRR5146865.15859310.1					Ш		C	TAA	CACCGCATCTAC
Frame 3					Ш			*	H R I Y

Physeter catodon conserved 1 nt deletion exon 2

SRA searched

SRR5146847- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 Sample ID: SAMN06187412 SRR5146843- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 Sample ID: SAMN06187413 SRR5146865- The Genome Center at Washington University School of Medicine in St. Louis 2017-01-05 Sample ID: SAMN06187411

	594	604	614	624 63	4	644	654
D+ BTA_NM_001206907.2 - EX2	GCCAGCGCCGGATAC	YČC ČČCČ	CGTGGTGCTCGTC	GCACTTCCTGCTGC	CC	GTGGCTGTTGTG	TGTTTCTGCTA
Frame 3	ASAGY	TAA		HFLL	E	VAVV	CFCY
D. COD MIN. 04.0070500 4-00002054 00004		100000000					
Frame 3	A S A R C	T V A	N V V V V	H F L P	E -	V A V V	C F C Y
FWD gnl SRA SRR5146865.23319104.1 Frame 3	GCCAG <mark>T</mark> GCCCGGT <mark>G</mark> C A S A <mark>R C</mark>	ACGG <mark>T</mark> GGC T V A	AGTGGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	C -	GIGGCIG V A	
FUD gnl SRA SRR5146843.13304771.1 Frame 3	GCCAG <mark>T</mark> GCCCGGTGC A S A R C	ACGG <mark>T</mark> GGC T V A	AGTGGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	E -	GIGGCIGIIGIG V A V V	TGCTTCTG TN A C F C
FUD gnl SRA SRR5146865.21925799.1 Frame 3	GCCAG <mark>T</mark> GCCCGGTGC A S A R C	ACGG <mark>T</mark> GGC T V A	AGTGGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	E -	GIGGCIGIIGIG V A V V	TGCTTCTGTA C F C Y
REV gnl SRA SRR5146865.6516907.1 Frame 3	GCCAG <mark>T</mark> GCCCGGT <mark>G</mark> C A S A R C	ACGG <mark>T</mark> GGC T V A	AGTGGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	C - E -	GIGGCIGIIGIG V A V V	TGCTTCTGTA C F C Y
REV gnl SRA SRR5146865.18729709.1 Frame 3	GCCAG <mark>T</mark> GCCCGGTGC A S A R C	ACGG <mark>T</mark> GGC T V A	AGTGGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	C - E -	GIGGCIGIIGIG V A V V	TGCTTCTGTA C F C Y
Fuo gnl SRA SRR5146847.7166341.1 Frame 3	GCCAG <mark>T</mark> GCCCGGTGC A S A R C	ACGG <mark>T</mark> GGC T <mark>V</mark> A	AGTGGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	C -	GIGGCIGIIGIG V A V V	TGCTTCTGTA C F C Y
REV gnl SRA SRR5146843.33083365.1 Frame 3	CCAG <mark>T</mark> GCC CGGTG C S A <mark>R C</mark>	ACGG <mark>T</mark> GGC T <mark>V</mark> A	AGTGGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	C -	GIGGCIGIIGIG V A V V	TGCTTCTGTA C F C Y
REV gnljSRAJSRR5146843.33070640.1 Frame 3	CCAG <mark>T</mark> GCCCCG <mark>GTG</mark> C S A R C	ACGG <mark>T</mark> GGC T V A	AGTGGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	E -	GIGGCIGIIGIG V A V V	TGCTTCTGTA C F C Y
FWD gnl SRA SRR5146843.26747385.2 Frame 3			GGTG <mark>G</mark> TCGT	CACTICCICCIC H F L P	10 - 2 -	GIGGCIGIIG <mark>G</mark> G V A V <mark>G</mark>	TGCTTCTG T TA C F C Y
REV gnl SRA SRR5146865.25177369.1 Frame 3			GIG <mark>G</mark> ICGI V V V	CACTICCICCIC H F L P	C -	GIGGCIGIIGIG V A V V	TGCTTCTGTA C F C Y
REV gnljSRAJSRR5146847.32187234.1 Frame 3				CACTICCICCIC H F L P	E -	GTGGCTGTTGTG V A V V	TGCTTCTGTTA C F C Y

Balaenoptera acutorostrata Conserved 1 nt deletion exon 2

SRA searched

SRR924087- Korea Institute of Ocean Science and Technology 2013-10-31 Sample ID: SAMN02192644 (MinkeWhale-01) SRR4011112-Institute of Marine Research 2016-08-13 Sample ID: SAMN05447714 (AT)

De BTA_NM_001206907.2 - EX2 Frame 3	NCGGCGCCCGCGCCGCCGCGCCGCGCGCGCGCGCGCGCG
FID NW_006725477.1:c560939-510360 Frame 3	TACGE GGCBCGCGCGCGCGCGCTCCCCCGCCGCCGCCGCCGCCGCCGCC
De FIID gnljSRAJSRR4011112.27205641.2	:ACGGTGGCAGGGGGGGGGGCGACCACTTCCTCCA
Frame 3	T V A V A V D H F L H
De FIID gnljSRAJSRR4011112.37032082.2 Frame 3	ACGGTGGCAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
De FIID gnljSRAJSRR4011112.77468341.2	:ACGGTGGCGGTGGGCGACCACTTCCTCCA
Frame 3	T V A V A V D H F L H P W
De FIID gnl SRA SRR924087.176242547.1	:ACGGTGGCGGTGGACCACTTCCTCCA-CCCNTGGCTGTTGTGT
Frame 3	T V A V A V D H F L H P W L L C
De REV gnl SRA SRR924087.45751921.2	:ACGGTGGCGGTGGCGGTCGACCACTTCCTCCA-
Frame 3	TVAVAVDHFLH-PWLLCA
De REV gnl SRA SRR924087.285524599.1	ACGGTGGCAGGGGGCGACCACTTCCTCCA
Frame 3	T V A V A V D H F L H P W L L C A S V
De FIIO gnijSRAjSRR924087.28626671.2	ACGGTGGCAGGGGGGGGGGCGACCACTTCCTCCA
Frame 3	T V A V A V D H F L H P W L L C A S V I C T
De Fillo gnijSRAjSRR4011112.22303842.1 Frame 3	:ACGGTGGCAGTGGCGGCGCGCCGCCCCCCCCCCCCCCCC
De REV gnl SRA SRR4011112.7606211.2	:ACGGTGGCAGTGGCGGCTCGACCACTTCCTCCACA-CCCATGGCTGTGTGTGTGTGTGTGTGTGTGTGTGCGCATCTGGGGGCGCTGGGGGCCTGGGGGC
Frame 3	T V A V A V D H F L H- P W L L C A S V I C A S G R W C
De REV gnl[SRA[SRR924087.331466944.1	:ACGGTGGCAGTGGCGGTCGACCACTTCCTCCA
Frame 3	T V A V A V D H F L H P W L L C A S V I C T S G R W C S
De REV gnl[SRA[SRR924087.134301300.1	:ACGGTGGCAGTGGCGGTCGACCACTTCCTCCA-CCCATGGCTGTTGGTTGGTTCTGTATCTGCACATCTGGGGGCTGGTGCTGC
Frame 3	T V A V A V D H F L H P W L L C A S V I C T S G R W C S
De REV gnl SRA SRR4011112.132497456.2	:ACGGTGGCAGTGGCGGTCGACCACTTCCTCCA-CCCATGGCTGTTGGTTGGTTCTGTTATCTGCGCATCTGGGGGCTGGTGCTGCGGG
Frame 3	T V A V A V D H F L H P W L L C A S V I C A S G R W C S G
De REV gnl SRA SRR4011112.132268527.1	:ACGGTGGCAGTGGGGGGTCGACCACTTCCTCCA-CCCATGGCTGTTGGTTGGTTCTGTTATCTGCGCATCTGGGGGCTGGTGCTGCGGG
Frame 3	T V A V A V D H F L H P W L L C A S V I C A S G R W C S G
De REV gnl SRA SRR924087.57482576.1 Frame 3	ACGGTGGCAGTGGCGGTCGACCACTTCCTCCACACCTCCACGGCTGTTGGTTG
De Feo gnijSRAjSRR4011112.51814560.2	CGGTGGCAGTGGCGCTCGACCACTTCCTCCACA-CCCATGGCTGTGTGTGTGTGTGTGTGTGTCTGTTATCTGCGCATCTGGGGCGCTGGTGCTGCGGG
Frame 3	V A V A V D H F L H- P W L L C A S V I C A S G R W C S G
De Feo gnijSRAjSRR924087.507482534.1	CAGTGGOGGTCGACCACTTCCTOCA-CCCATGGCTGTTGGTTGGTTCTGGTCTGCACATCTGGGOGCTGGTGCTGCGGG
Frame 3	V A V D H F L H P W L L C A S V I C T S G R W C S G
Example 3	CGGTCGACCACTTCCTCCA-CCCATGGCTGTTGGTTGCTTCTGCTGCTGCATCTGGGCGCCTGGTGCTGCGGG D H F L H P W L L C A S VII C A S G R W C S G
De Feo gnijSRAJSRR4011112.100241100.2	GACCACTTCCTCCACA-CCCATGGCTGTTGTGTGTGTGTCTGTTAMCTGCGCATCTGGGGGCTGGTGCTGCGGG
Frame 3	D H F L H- P W L L C A S VII C A S G R W C S G
De REV gnl SRA SRR4011112.166394406.1	CACTTCCTCCCCCATGGCTGTTGTGTGTGTGTGTGTTCTGTTATCTGCGCATCTGGGGGCTGGTGCTGCGGG
Frame 3	H F L E- P W L L C A S VII C A S G R W C S G
De REV gnl SRA SRR924087.144891681.1	ACTICCIÚCA-CCCAIGGCIGIIGGUITCIGUITCIGUITCIGUACAICIGGGUGCIGGIGCIÚCGGG
Frame 3	F L <mark>H</mark> P W L L C A S VII C I S G R W C S G
De REV gnl SRA SRR4011112.112133539.2	CM-CCCMTGGCTGTTGTGTGTGTGTGTGTTTCTGTTAMCTGCGCATCTGGGGGCCTGGTGCTGCGGG
Frame 3	H_ P W L L C A S V I C A S G R W C S G
De Fillo gnijSRAjSRR924087.418239962.1	GGCTGTTGTGTGTGTGTGTGTGCTGCACATCTGGGGGGCTGGTGCTGCGGGG

Balaenoptera bonaerensis Conserved 1 nt deletion exon 2 SRA searched

SRR4011114- Institute of Marine Research 2016-08-13 Sample ID: SAMN05447715 SRR4011113- Institute of Marine Research 2016-08-13 Sample ID: SAMN05447715

	00 610	620	630	640	650	660	670	680	690
De BTA_NM_001206907.2 - EX2	TACGGCGGCCGT	GGTGCTCGTGCAC	TTCCTGCT	CCCGTGGCTG	TTGTGTGTT	TETGETACCTGC.	ACATCIGGGIG	CTGGTGCTGCC	GTGCCC
Traine I									
REV DF429831.1 Frame 1	CACGGIGGCAGI T V A V	GGCGGGTCGCAC	TTCCTCCC F L P	CCCATGGCTG P M A	TTG <mark>C</mark> GNNNN V A ?	NNNNNNNNNNNN ? ? ? ?	<u>NNNNNNNN</u> TG ?	CTGGTGCTCC L V L	G <mark>G</mark> GCC <mark></mark> R A
FID gnl SRA SRR4011113.63527837.1 Frame 1	CACGGTGGCAGT TVAV	GGCGGTTGTCCAC	TTCCTCC F L P	CCCATGGCTG P M A	Ţ				
REV gnl SRA SRR4011113.63527837.2 Frame 1	CACGGTGGCAGT TVAV	GGCGGTTGTCCAC	F L P	CCCATGGCTG P M A	TTGCGTGCT V A C	TCTGTTACCTGC F C Y L	GCATCTG R I		
FID gnl SRA SRR4011114.65796625.2 Frame 1	CACGGTGGCAGT TVAV	GGCGGTTGTCCAC	TTCCTCC F L P	CCCATGGCTG P M A	TTGCGTGCT V A C	TCTGTTACCTGC F C Y L	GCATCTGGG R I W		
FID gnl[SRA[SRR4011114.63475535.1 Frame 1	CACGGTGGCAGT TVAV	GGCGGTTGTCCAC	F L P	CCCATGGCTG P M A	TTGCGTGCT V A C	TCTGTTACCTGC F C Y L	GCATCIGGG R I W		
FID gnl[SRA[SRR4011113.205833559.1 Frame 1	CACGGTGGCAGT TVAV	GGCGGTTGTCCAC	F L P	CCCATGGCTG P M A	TTGCGTGCT V A C	TCTGTTACCTGC F C Y L	GCATCIGGGIG R I W V	C	
REV gnl SRA SRR4011113.44932454.1 Frame 1	CACGGTGGCAGT TVAV	GGCGGTTGTCCAC	TTCCTCC F L P	CCCATGGCTG P M A	TTGCGTGCT V A C	TCTGTTACCTGC F C Y L	GCATCIGGGIG R I W V	CTGGTGCTCC L V L	G <mark>G</mark> GCC <mark>I</mark> R A
FID gnl SRA SRR4011114.122755535.2 Frame 1	CACGGTGGCAGT TVAV	GGCGGTTGTCCAC	F L P	CCCATGGCTG P M A	TTGCGTGCT V A C	TCTGTTACCTGC F C Y L	GCATCIGGGIG R I W V	CTGGTGCTCC L V L I	G <mark>G</mark> GCC
FID gnl[SRA[SRR4011113.143169.2 Frame 1			L	CCCATGGCTG P M A	TTGTGTG C T V V C	TCTGTACCTGC F C Y L	GCATCTGGGCG R I W A	CTGGTGCTCC L V L	G <mark>G</mark> GCC <mark>I</mark> R A
FID gnl SRA SRR4011114.110600766.1 Frame 1				CCATGGCTG M A	TTGTGTG C T VVC	TCTG TAT CTGC F C Y L	GCATCTGGGCG R I W A	CTGGTGCTCC L V L	G <mark>G</mark> GCC <mark>A</mark> R A
FID gnl SRA SRR4011114.54753601.1 Frame 1				CATGGCTG M A	TTGCGTGCT V A C	TCTGTTACCTGC F C Y L	GCATCIGGGIG R I W V	CTGGTGCTCC L V L I	G <mark>G</mark> GCC
FID gnl[SRA[SRR4011113.68591068.2 Frame 1				G	TTGCGTGCT V A C	TCTGTTACCTGC F C Y L	GCATCTGGGTG R I W V	CTGGTGCTCC L V L	G <mark>G</mark> GCC <mark>I</mark> R A
FID gnl SRA SRR4011113.101710613.1 Frame 1				G	TTGTGTG C T V V C	TCTGTTACCTGC F C Y L	GCATCTGGGCG R I W A	CIGGIGCICC L V L	G <mark>G</mark> GCC R A
FID gnl[SRA[SRR4011114.157540768.2 Frame 1					GTGTG C T V C	TCTG T TA T CTGC F C Y L	GCATCTGGGCG R I W A	CTGGTGCTCCG	g <mark>g</mark> gcc <mark>e</mark> R A

Balaena mysticetus Conserved 1 nt deletion exon 2

SRA searched

SRR1685385- University of Liverpool 2015-01-06 Sample ID: SAMN03225705 (Bowhead ID325) SRR1685386- University of Liverpool 2015-01-06 Sample ID: SAMN03225705 (Bowhead ID325)

C+ BTA NM 001206907.2 - EX2	SACGCCAGEG	CCGGATA	CACGGE	GCCGT	GETEET	GTGE	ACTT	CCTG	TG	cee	GIGG	CTGT	TGT	TGT	TTCT	GCT	ACCT	GCA	ATC	TGGG
Frame 2	TAS	A G Y	TA	A V	VL	V	H F	L	I	P	V	AV	V	C	F	C	Y III	H	Ι	W
De REV gnljBL_ORD_ID[1482 Frame 2	GACGGCCAG	CCCGGTA A R Y	CACGG	GGCAGT A V	GGCGGT A V	CGTOC V	ACTT H F	CCTC	210 P	CCC P	ATGG M	CTGT A V	TGT	GICI GICI R	TTCT	G T I	ACCI Y I	GC <mark>G</mark> R	ATC I	TGGG
De Feo gnijSRAjSRR1685385.291823986.2 Frame 2	GACGGCCAGTG	CCCCGGTA A R Y	CACGGT T	GGCNGT A V	GGCGGT AV	CGTCC V	ACTT H F	CCTC	- <mark>1</mark>	2										5,415
De Feo gnijSRAjSRR1685385.219098232.1 Frame 2	SACGGCCAG ^{II} G T A S	CCOGGTA A R Y	CACGGT T	A V	GGCGGT AV	CGTOC V	ACTT H F	CCTC	- E	PCC	MTGG	C								
De Feo gnijSRAJSRR1685385.191166236.1 Frame 2	GEG	CCCGGTA A R Y	CACGGT T	A V	A V	CGTCC V	ACTT H F	CCTC	2 - P -	PCCC	MTGG	CTGT A V	TGT	R	TTCT	GTI	ACCI Y I	GCG	IATC	TGGG
De Fill gnljSRAJSRR1685385.257870608.1 Frame 2	G	CCCCGTA A R Y	CACGGT T	A V	A V	V	ACTT H F	CCTC	- E	PCC	MTGG	CTGT A V	TGT	R	TTCT	GTT	ACC1	GCG	CATC I	TGGG
De FIID gnl SRA SRR1685385.105117927.2 Frame 2		COGGTA R Y	CACGGT T	A V	A V	V	ACTT H F	CCTC	- D -	CCC P	ATGG	CTGT A V	TGT	R	TTCT	GTT.	ACCI Y I	GCG	CATC I	TGGG
De Fill gnljSRAJSRR1685385.127716020.1 Frame 2		COGGTA R Y	CACGGT T	A V	A V	V	ACTT H F	CCTC	20 - P -	P	MTGG	CTGT A V	TGT	R	TTCT	GTT.	ACCI Y I	GCG	ATC I	TGGG
De Fill gnijSRAJSRR1685385.125328155.1 Frame 2		COGGTA R Y	CACGGT T	A V	G G G T	CGTCC V	ACTT H F	CCTCC	- <mark>10</mark>	P	MTGG	CTGT A V	TGT	R	F	YT	ACCI Y I	GCG	CATC I	TGGG
De Fillo gnijSRAJSRR1685385.86400789.1 Frame 2		COGGTA R Y	CACGGI	A V	A V	CGTCC V	ACTT H F	CCTC	2 - P -	PCC	ATGG	CTGT A V	TGT	R	TTCT	GTT.	ACCI Y I	GCG	ATC I	TGGG
De Fill gnljSRAJSRR1685385.293698380.1 Frame 2					AV	V	ACTT H F	CCTC	- <mark>-</mark>	PCC	MTGG	CTGT A V	TGT	R	TTCT	GTI	ACCI Y I	GCG	IATC	TGGG
De FIID gnl SRA SRR1685385.251121772.1 Frame 2					GGCGNT A ?	CGTCC V	ACTT H F	CCTC	- E	PCCC	MTGG	CTGT A V	TGT	R	TTCT	G T T	ACCI I I	GCG	ATC	TGGG
De Feb gnl SRA SRR1685385.96190329.1 Frame 2					GEGGT	CGTCC V	ACTT H F	CCTC L	- <mark>0</mark>	P	ATGG	CTGT A V	TGT	R	TTCT	GTT, C	ACCI Y I	GCG	ATC I	TGGG
De Fill gnijSRAJSRR1685385.124441553.2 Frame 2					GCGGT A V	V	ACTT H F	CCTC L	- <mark>D</mark>	P	MTGG	CTGT A V	TGT	R	TTCT	GTT.	ACCI Y I	GCG	ATC	TGGG
De Fill gnl SRA SRR1685386.350533959.2 Frame 2					GGT	V	ACTT H F	CCTC	20 P	P	MTGG	CTGT A V	TGT	R	TTCT	GTT	ACCI Y I	GCG	ATC	TGGG
De Feo gnijSRAjSRR1685385.175143310.2 Frame 2									ю ₋	P C C C	MTGG	CTGT A V	TGT	R	F	G T I	ACCI Y I	GCG	I ATC	TGGG
De REV gnijSRAJSRR1685385.151491640.2 Frame 2													GT	R	TTCT	G T I C	ACCI Y I	GCG	ATC	TGGG

Figure S19: SRA validation for conserved nucleotide deletion in exon 2 of *Mtnr1b* from *D. leucas*, *N. asiaeorientalis*, *P. catodon*, *B. acutorostrata*, *B. bonaerensis* and *B. mysticetus*.

Eschrichtius robustus SRA reads bridging Exon 2 282bp deletion (NNN) SRA searched

SRR5495100- Purdue University 2017-05-02 Sample ID: SAMN06837694 (GFD-02) SRR5495106- Purdue University 2017-05-02 Sample ID: SAMN06837692 (ER-14-168)

	208	218	228	238	2.48	258	268	278	288	298	3	908	318	
D* BTA_NM_001206907.2 - EX2	TGCCGCAGCG	GACCTACCAC	CGCCTCTG	ccecceccec	ACGCCGCCG	CTACGTTNNNT	GGTCTAG	TOGOGTGCGGAG	CTTTCTGA	GCATGTTTS	TGGTCTT	TGTGATC	TTCGCCAT	ICT.
Framer	CKO		K B C				N 0 0	K K C						
	2													~
Frame 1	C H S	IGACCTACCAC / T Y H	R I Y	C H W -					- F L	T T F	TGGT <mark>G</mark> TT M V F	TGTGATC V I	F V J	I I
REV gnl[SRA[SRR5495106.39817980.1 Frame 1	TGCCACAGCG	NGACCTACCAC / T Y H	R I Y	CH W -					-TTTCTGA - F L					
REV gnljSRAJSRR5495100.66934205.1 Frame 1	TGCCGCAGCG	NGACC <mark>G</mark> ACCAC	CGCATCIG R I C	C R W -					-TTTCTGG	CACGTTCA A T F	1			
REV gnIJSRAJSRR5495106.28485631.1 Frame 1	C H S	NGACCTACCAC	R I Y	CHCCACOGG-					-TTTCTGA	CACGTTCA T T F	1			
Fito gnl[SRA[SRR5495106.29362269.1 Frame 1	TGCCAGCG	GACCTACCAC	CGCATCTA R I Y	CHCCACCGG-					-TTTCTGA	CACGTTCA T T F	tr.			
REV gnl[SRA[SRR5495100.94761215.2 Frame 1	TGCCACAGCG	NGACCTACCAC	CGCATCTA R I Y	CHCCACOGG-					-TTTCTGA	CCACGTTCA T T F	TGGT			
Find gnl[SRA[SRR5495100.177319358.2 Frame 1	TGCCACAGCG	MACCTACCAC	CGCATCTA R I Y	C H W -					-TTTCTGA	CCACGTTCA T T F	TGGT <mark>G</mark> TT N V F	TGTGATC	TTCGCAI	T
REV gnl[SRA[SRR5495100.372617229.2 Frame 1	TGCCACAGCG	NGACCTACCAC	CGCATCTA R I Y	CHW-					-TTTCTGA - F L	CACGTTCA T T F	TGGT <mark>G</mark> TT N V F	TGTGATC V I	TTCGCAI	TCT
REV gnl[SRA[SRR5495106.90703659.2 Frame 1	TGCCACAGCG	NGACCTACCAC	CGCATCTA R I Y	C H W -					-TTTCTGA	CCACGTTCA T T F	TGGT <mark>G</mark> TT N V F	TGTGATC	TTCGCAI	TCT I
REV gnl[SRA[SRR5495100.35089126.2 Frame 1	TGCCACAGCG	GACCTACCAC	CGCATCTA R I Y	C H W -					-TTTCTGA - F L	CACGTTCA T T F	TGGT <mark>G</mark> TT NV F	TGTGATC V I	TTCGCAT	TCT
REV gnljSRAJSRR5495100.57780103.2 Frame 1	TGCCACAGCG	NGACCTACCAC	CGCATCTA R I Y	CH W -					-TTTCTGA	CACGTTCA T T F	TGGT <mark>G</mark> TT N V F	TGTGATC V I	TTCGCAT	TCT
Find gnl[SRA]SRR5495106.19866462.1 Frame 1	C H S	NGACCTACCAC / T Y H	CGCATCTA R I Y	C H W -					-TTTCTGA - F L	CACGTTCA T T F	TGGT <mark>G</mark> TT MV F	TGTGATC V I	TTCGCAI	TCT I
Filo gnl[SRA[SRR5495100.189898729.1 Frame 1	TGCCACAGCG	NGACCTACCAC	CGCATCTA R I Y	CHW-					-TTTCTGA - F L	CACGTTC T T F	TGGTGTT L V F	TGTGATC V I	TTCGCAI	TCT I
Fill gnl[SRA[SRR5495106.82573747.2 Frame 1	TGCCACAGCG	NGACCTACCAC	CGCATCTA R I Y	CHCCACCGG-					-TTTCTGA	CCACGTTCA T T F	TGGT <mark>G</mark> TT M V F	TGTGATC V I	TTCGCAT	TCT I
Fill gnl[SRA[SRR5495106.71475325.1 Frame 1	TGCCACAGCG	NGACCTACCAC / T Y H	R I Y	CHCCACOGG-					-TTTCTGA	CCACGTTCA T T F	TGGT <mark>G</mark> TT MV F	TGTGATC V I	TTCGCAT	TCT I
REV gnl[SRA[SRR5495106.81848423.1 Frame 1	TG TATA AG A G	S Y H	R I Y	CHCCACCGG-					-TTTCTGA	CACGTTCA T T F	TGGT <mark>G</mark> TT N V F	TGTGATC V I	TTCGCAT	TCT I
Fillo gnl[SRA[SRR5495106.71540003.2 Frame 1	TGCCACAGCG	NGACCTACCAC / T Y H	CGCATCTA R I Y	CHW-					-TTTCTGA - F L	CCACGTTCA T T F	TGGT <mark>G</mark> TT N V F	TGTGATC V I	TTCGCAI	TCT I
Fill gnl[SRA[SRR5495100.79781509.2 Frame 1	TGCCAGCG	GACCTACCAC	R I Y	CRCCACRG-					-TTTCTGA - F L	CACGTICA T T F	TGGT <mark>G</mark> TT MV F	TGTGATC V I	TTCGCAT	TCT I
Feo gnl[SRA[SRR5495100.151215035.2 Frame 1	C H S	GACCTACCAC 7 T Y H	R I Y	CHGCCACOGG-					-TTTCTGA	CACGTTCA T T F	TGGT <mark>G</mark> TT	TGTGATC V I	TTCGCAT	ICT.
Fino gnl[SRA]SRR5495106.78827869.2 Frame 1	CG	GACCTACCAC	R I Y	CH W-					-TTTCTGA	CACGTTCA T T F	TGGT <mark>G</mark> TT	TGTGATC	TTCGCAT	ICT I
REV gnIJSRAJSRR5495100.290411448.1 Frame 1		TACCAC Y H	R TCTA	CHCCACCGG-					-TTTCTGA	CACGTTCA T T F	TGGT <mark>G</mark> TT	TGTGATC V I	TTCGCAT	TCT I
Find gnl[SRA[SRR5495100.322018798.2 Frame 1		CCAC	CGCATCTA R I Y	CH W -					-TTTCTGA	CCACGTTCA T T F	TGGT <mark>G</mark> TT M V F	TGTGATC V I	TTCGCAT	TCT I
Feo gnl SRA SRR5495100.295062850.2 Frame 1		AC	R I Y	CH W -					-TTTCTGA - F L	CACGTTCA T T F	TGGT <mark>G</mark> TT	TGTGATC V I	TTCGCAI	TCT I
REV gnl SRA SRR5495100.465289704.1 Frame 1			CGTCTA	CH W -					-TTTCTGA - F L	TCACGTTCA	TGGT <mark>G</mark> TT	TGTGATC V I	TTCGCAT	TCT I
REV gnl SRA SRR5495100.420724110.1 Frame 1				C H W -					-TTTCTGA - F L	CACGTICA T T F	TGGT <mark>G</mark> TT N V F	TGTGATC V I	TTCGCAT	FCT I

Figure S20: SRA validation for nucleotide deletion in exon 2 of *Mtnr1b* from *E. robustus*.

Hippopotamus amphibius Asmt Exon 1 assembly using SRA

SRA searched

WGS

SRR5663648-Senckenberg Gesellschaft fuer Naturforschung **Released:** 2018-04-05 **Sample:** SAMN07206988 SRR5663647-Senckenberg Gesellschaft fuer Naturforschung **Released:** 2018-04-05 **Sample:** SAMN07206988 **RNA-seq**

SRX1164570-SRR2183469 School of Biological & Chemical Sciences, Queen Mary University of London Released: 2015-09-30 Sample: SAMN04009372

C+ HSA NM_004043.2 Frame 1	AT SCARTON CACAGE CONSCIENCE ACTACE AND A CONTRACTACE AND A CONTRA
De FIID gnijSRAjSRR2183469.27421052.1	
Ce REV gnl SRA SRR5663647.40717428.2	CONSTRUCTION OF A CONSTRUCTICA CONSTRUCTION OF A
C+ F#D gnl SRA SRR5663648.53806247.2	STGGCAGGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGACTGCCGTCTGCTGAGGGAATATCCCAGTGGCTTC
De REV gnl SRA SRR5663647.251145673.2	STGGCAGGAAGATGGGCTCCTCAGGGGGGGGGGGGGGGGG
C+ REV gnl SRA SRR5663647.64973964.1	STGGENGGANGATGAGETCCTCAGAGGAGGAGGAGGAGGAGGAGGAGGGAGGGGGCTTCA
Ce REV gnljSRAJSRR5663647.35346080.2	TEGERAGARGATEGECTCCTCAGAGGAGGAGGAGGACTGCCGTCTGCCGCAATATTCCCACTGCCTCAT
De REV gnl SRA SRR5663647.186096856.2	STGGCAGGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGAGGAGGGCTGCCGGCGGAGGGGCTCCATG
Ce FID gnl SRA SRR5663647.212043975.2	STGGCAAGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGAGGACTGCCGTCTGCCGGCAATATTCCCAGGGCTTCATGG
C+ REV gnl SRA SRR5663647.447858890.1	STGGCAGGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGAGGACTGCCGCCGCCGCCGCAGGGAATATTCCCAGGGCTTCATGGCC
De REV gnijSRAjSRR5663647.17874819.1	STGGUAGGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGACTGCTGTCGTGAGGAATATCCCAGGGCTCATGGTC
C+ FID gnl SRA SRR5663648.120212257.1	STGGCAGGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGAGGACTGCTGGCGGCATGGCGAGTACTGGCGTCCTCGAGGGCTTCCTGGAGT
C+ FID gnl[SRA[SRR5663648.14385048.1	STGGUAGGAAGATGGGUTCCTCAGAGGAGGAGGAGGAGGACTGUCGTCTGAGGGAATATCCAGGGGCTTCATGGTUT
C+ REV gnl SRA SRR5663647.248213150.2 Frame 1	SUGGUAGGAAGATGGGUTCCTCAGAGGAGGAGGAGGAGGACTGUCGCTGAGGGAATATCCCAGGGCTTCATGGTUT
C+ FIP gnl[SRA[SRR5663648.69947600.1	STIGGEAGGAAGATGGGETCCTCAGAGGAGGAGGAGGAGGAGGAGGGGGATATTCCAGGGGCTTCATGGGET
C+ REV gnl SRA SRR5663647.255385547.2	STGGEARGAAGATGGGETCCTCAGAGGAGGAGGAGGAGGAGGAGGGGGAGGGGGGCTCCATGGCTTCGGGCTTCATGGCTTCG
C+ REV gnl[SRA[SRR5663647.354510019.1	STGGCAGGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGAGGCGCCGCCGCGCGCG
C+ REV gnl SRA SRR5663648.17974644.2	STGGCAGGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGAGGAGGGGAGGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCAGGCTCCATGGCTCCCCCCCC
Ce FID gnl[SRA[SRR5663647.347816954.1	STGGEAGGAAGATGGGETCCTCAGAGGAGGAGGAGGAGGAGGAGGGGGAGGGGGCTCCTCGGCTCCTCGGGTCCCCAGGGGCTCCTCGCGGCTCCTCGCGCGCCCGCGCGCCCGCGCGCCCGCGCGCCCGCGCGCCCGCCGCG
C+ REV gnllSRAJSRR5663648.93316220.2 Frame 1	STGGUAGGAAGATGGGUTCCTCAGAGGAGAGAGGAGGAGGAGGGGAGTGGCGGGAGTATTCCAGGGUTCGCAGGTUTCGCUTCGC
De Feo gnijSRA(SRR5663647.205629852.1 Frame 1	STGGCAGGAAGATGGGCTTCCTCAGAGGAGGAGGAGGAGGAGGAGGGCTGCTGGCGGCTTCATGGCTTCCGCAGGTA
Ce REV gnl[SRA[SRR5663647.116173870.2 Frame 1	STGGERAGGARGATGGGETCCTCAGAGGAGGAGGRGGRGGRGGCGECTGCTGAGGGAATATCCAGEGGCTTCATGGTETCGCAGGTAG
De REV gnl SRA SRR5663648.97953428.1	TEGETAGGAAGATGGGETCCTCAGAGGAGGAGGAGGAGGAGGAGGGGGGAGGAGGGGGCTCCATGGTTCGCAGGTAGGA C C C C C C C C C C C C C C C C C C C
C+ REV gnl SRA SRR5663648.88376996.2 Frame 1	STGGCAGGAAGATGGGCTCCTCAGAGGAGGAGGAGGAGGAGGGCTGCTGCGGGCTTCATGGCTCCTCGCAGGTAGGA G R R M G S S E E E E D C R L L R E Y S S G F M V S O V G
C+ REV gnl[SRA[SRR5663648.63647193.1 Frame 1	TEGERAGGARGATGGGETCCTCAGAGGAGGAGGAGGAGGAGGAGGGGGGGGGG
C+ REV gnl SRA SRR5663648.52634068.2 Frame 1	STGGUAGGAAGATGGGUTCCTCAGAGGAGGAGGAGGAGGACGUCGUCGUCGAGGGAATATCCCAGGGCTTCATGGTUTCGCAGGAGGA
De REV gnl[SRA[SRR5663648.167799331.2	JERGERAGARGEGEGCTCCAGAGGAGGAGGAGGAGGACGCCCCCGCGGAGAGARGEGCTCCAGGGCTCCAGGGCAGGA

Red box highlights a spliced RNA-Seq read spanning from exon 1 to exon 2

Hippopotamus amphibius Asmt Exon 5 assembly using SRA

SRA searched

WGS

SRR5663648 -Senckenberg Gesellschaft fuer Naturforschung Released: 2018-04-05 Sample: SAMN07206988

SRR5663647 - Senckenberg Gesellschaft fuer Naturforschung Released: 2018-04-05 Sample: SAMN07206988

SRR5663646- Senckenberg Gesellschaft fuer Naturforschung Released: 2018-04-05 Sample: SAMN07206988

RNA-seq

SRX1164570-SRR2183469 School of Biological & Chemical Sciences, Queen Mary University of London Released: 2015-09-30 Sample: SAMN04009372

Frame 3	LARAGETTETEROVICEATERACHIOTOCHOSOCIADOSCERCIOSTECENCIADOSTECE
Per gnijskajski 5663648.15688563.1	UICCUARDOCALBURCEW BUICLEVED OBCIDE DU ANTE DUAL DE L'ORANDO CALDU DU CENTRA DE L'ELTRA CEDITERNO DE LA CALDU A
Frame 3	
Feb gnl[SRA[SRR5663648.120489597.1	
File gnt(SRA(SRR5663648.14487406.1 Frame 3	GTCCGAGGGGGARAGCTEGESTCCATALGEGGEGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Pro gnijSRAJSRR5663648.157661212.1	STCCAAGGCCAARGCTBCCBCTBCATCCGBCGCTBCCAGCABCTGCAGGCGBCACGTBCCTBCACCTSCCTBCCCTSCCTBCCCTSCCBCCBACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
Fe0 gnl(SRA(SRR5663648.143654701.1 Frame 3	OTCCAAGGGCGAAGGCCGCGCGCGCGCGCGCGGCGGCGGCGG
De Feo gnijSRAjSRR5663646.221767122.2	GTCCAAGGCCAREGCTBCGCTFCATCCGG-GBCTSCAGGCTSCAGGCGBCAAGGTSGAGGTGCAGGCGCTSCCCTBCACCTSCGGGGCTCCCA S E G E E L ERIE N EN ROSL O DE V N E E V L V A F D L S F F
De Feo gnijSRAjSRR2183469.38701929.1 Frame 3	CGARGEGGARAGETEGENTERATIONS/HECTRICADADESTENSAMONDERCOMARGODEARGETOETGENESTENSECTETTGARCEDETG E G E R. 1961 F MIRTON 10 D V MIRTON 6 R EM V 1 AA P D 1 3
De Filo gnijSRAjSRR5663647.249650964.2	CRARGECCRAREGETECCEACOUTECCEACEUS DE CONTRECCEACEUR DE CONTRECCE
De Hev gnijSRAjSRR5663647.280703215.2 Frame 3	COARGO COARGO CONTRATICATION COMPANIANT COARGO AND TRAVIS DE BARGITACTICO CONTRACTICO CONT
De Filo gnijSRAjSRR5663647.141238597.1 Frame 3	AGGOCIALBUCTO CONTENTIONS CONSIGNATION CONTRACTOR CONTRACTOR CONTENTIAL
De Filo gnijSRAJSRR5663646.120725903.1	GORGENING OF CONTENTED CONCERCINGER DE CONCERCIÓN DE LA CONTENCIÓN DE LA C
D= FRV gnl(SRA(SRR5663647.184324525.2 Frame 3	GREGENISSICE GENERATER DE CONCARRA DE TERMENSION DE LA SECTETA DE CONTRA CENTRE CONCARRA CONTRA CONTRA CENTRE CONCARRA DE LA CONTRA CENTRE CONTRA CONTRA CENTRE
De fev gnijSRAjSRR5663647.95822490.2 Frame 3	George Aussie Constructions of Construction and Statemark and Constructions and Construction and Constructio
C+ Feb gnl[SRA[SRR5663647.279443429.1 Frame 3	SCALARSOCTEC, SCATTERING GERGECTSCARSOLARSTERING ROBERING GERANG GERCHSCHTTAACTSTOGERTTCCCARGEATERS E B L RF Y M R GEL Q DE V M R LEGEL Q A F D L S FF F F R R
De Fe0 gnl[SRA[SRR5663647,6029619.2 Frame 3	GCALARGEGE GOTTELATOGOGIGEGEGEGAGADETEGEARGEGEGARGETEGERAGETEGETEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEG
De Feo gnijSRAjSRR5663647.263135732.1 Frame 3	CARAGEGE COTTACTSCOB.GCCTSCASGADETGOSAGEGEGAGETGOSAGETGOSAGETGOSAGETGOSAGETTCCCASGATETSC E R L ME F M REGEL O UDE V M RELAGE G R REV L LA A P D L S PL F P REC C
De 589 gni[SRA[SRR5663646.272782468.2 Frame 3	CAN BRIGE CONTENTION OF CONCRECTION AND TO CONFIDENCE THAT TO A THE CONTENT OF CONTENTION OF THE CONTENT OF CONTENT.
De fev gnljSRAjSRR5663647.64393322.2 Frame 3	CGARANGET BEGET TELEVISED CONCECTUE AGARANGET CALSEGO BE BALGET SECTOR DE DECTIVALET TELEVISED E R L REF N R GOL O DO V N R LELE G R R V L A F D S P P P R C C
Co Hev gnl[SRA[SRR5663646.208845778.1 Frame 3	GALEGAC GALEGA
De Filo gnijSRAjSRR5663647.420108458.2 Frame 3	GARANGET BEGENTER TO CONCOME TO CASE AND THE OFFICE A REPORT OF A REPORT OF A REPORT OF A CONCEPT OF A CONCEP
De Rev gnl[SRA[SRR5663646.89299604.2 Frame 3	A GOLTE CATE CONCECTENCAS CALLER AND TENDES AND TO CONTRACT TO CONTRACT STOCED TO CONCERT TO CONCERT STACE R LINES IN RING LO DE V N REAL STACE S R V LIA A PD LS PP P VALOR CONTRACTS CONCERT CONCERTS SACC
C+ Fe0 gnl[SRA[SRR5663647.344730297.1 Frame 3	A SOCTES CONTEXTOC DO CONCERCIONAL DE CONCERCIONAL DE CONTEXTO CONTEXTO CONCENTRACENCIA CONTEXTO DE ACCENTENCE A CONTEXTO DE ACCENTENCE A CONTEXTO DE ACCENTENCE A CONTEXTO DE ACCENTENCE A CONTEXTO DE ACCENTEXTO D
De fev gnl SRA SRR5663647.332314222.2 Frame 3	GCTGGGGTTGCASGGGGCTGCASGAADCTGCASGGGCGCAAGGTGCASGGGCGCGGAGGTGCCAAGGATGCCGCCTTAACCTSTCGCGGTTCCCAAGGATGTSGAACCTT L GG F M R LGG L O DD V M R LGG L V L A F D L S FP F P R LG C D L
De Hev gnl(SRA(SRR5663647.354052102.2 Frame 3	GETERGETERADE GE GERGEGERGADE AD TERGEGERGADE AD GETERGEDE DE GETERGETERGEGETERGEGERGEGERGEGERGEDE ACET I DE TERMINE GEL DIDE VIER DE GERGERGEDE DE
De Fato gnl[SRA[SRR5663647.334725236.2 Frame 3	TO CONTRACTOR OF OF CONTRACTOR AND TO THE ADDITION AND ADDITION AND ADDITION ADDITION ADDITIONAL ADDITION ADDITIONAL
Rev gnijSRAJSRR5663646.53240602.1 Frame 3	TICATH: OB: GBCTRCARSAD TRONS AND FOR AN ODD BASES THE THE ADDRETT ACCEPT CONCENT CARDENT CONSTRUCTION AND THE CONSTRUCTION AND ADDRETT ADDRET
Fito gnijSRAJSRR5663647.238136676.1 Frame 3	TTCAPP.00-00-TRACEADEDTOSIAGETCALAGETCALAGEDICTOSACEDICTTAACCEDICGECTCCCACGGATERDEBACCTTOPA F N RE GL D D V N RE LA G S REV L ALA F D L S F F P R RE C D L G

Red box highlights RNA-Seq read mapping to exon 5

Figure S21: H. amphibius Asmt exon 1 and 5 assembly.



Figure S22: Maximum likelihood phylogenetic analysis of Perissodactyla, Artiodactyla and Primate *Aanat* genes. Node values represent branch support using the aBayes algorithm. In red the clade containing the *H. amphibius*.

0.01



Figure S23: Maximum likelihood phylogenetic analysis of Perissodactyla, Artiodactyla and Primate *Mtnr1a* genes. Node values represent branch support using the aBayes algorithm. In red the clade containing the *H. amphibius*.

H-



Figure S24: Maximum likelihood phylogenetic analysis of Perissodactyla, Artiodactyla and Primate *Mtnr1b* genes. Node values represent branch support using the aBayes algorithm. In red the clade containing the *H. amphibius*.

-



Figure S25: *T. manatus latirostris* gene annotations. Schematic representation of the gene structure of human and B. taurus *Aanat* (A), *Asmt* (B), *Mtnr1a* (C) and *Mtnr1b* (D) genes, each box represents an exon and lines represent intronic region, and sSchematic representation of the corresponding genes identified in *T. manatus latirostris* and location of the identified mutations. Non canonical splice sites are indicated above the corresponding annotation.

Trichechus manatus latirostris

SRA searched: SRR331137 - Broad Institute (BI) 2011-08-17 **Sample ID:** SAMN00632092 SRR331138 - Broad Institute (BI) 2011-08-17 **Sample ID:** SAMN00632092 SRR331139 - Broad Institute (BI) 2011-08-17 **Sample ID:** SAMN00632092 SRR331142 - Broad Institute (BI) 2011-08-17 **Sample ID:** SAMN00632092

BTA Aaanat - ex1 Frame 1	ATGTCCACGCCGAGCATCCACTGCCTGAAACCCTCGCCTCTGCACCTGCCGTCT
Rev gi 460718331 ref NW_004443964.1 :5926461-6165714 extraction	CCAGGACGTCCACGCAGAGAG <mark>IG</mark> TCCACCCCITGACGCCIGCACCCTGIACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
F00 gnl SRA SRR331137.208480462.3 G Frame 1	CCAGCACCTCCACGCAGAGAGTGTCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCC A R T S T Q S V H P L T P A T L Y L P
FuD gnl SRA SRR331139.155222636.1 G	CCAGCACCTCCACGCAGAGAGTGTCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCC A R T S T Q S V H P L T P A T L Y L P
Fu0 gnl SRA SRR331142.213243558.1 G	CCAGGACGTCCACGCAGAGAG <mark>IG</mark> TCCACCCTTGACGCCTGCACCCTGTACCTGCCGCC A R T S T Q S V H P L T P A T L Y L P
REV gnl SRA SRR331138.8422870.3 Frame 1	CCAGGACGTCCACGCAGAGAGAGAGTGCCACCCTGCACCCCGCACCCCGCCACCCCCGCACCCCCGCACCCCCC
REV gnl SRA SRR331137.194184985.1 G Frame 1	CCAGGACCTCCACGCAGAGAG <mark>TG</mark> TCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331139.70827001.1	CCAGGACGTCCACGCAGAGAG <mark>IG</mark> TCCACCCTTGACGCCACCCTC <mark>T</mark> ACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331137.121569433.1 G	CCAGGACCTCCACGCAGAGAG <mark>TG</mark> TCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331139.199392699.1 G Frame 1	CCAGGACCTCCACGCAGAGAG <mark>TG</mark> TCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
REV gnl SRA SRR331142.187479798.3 Frame 1	CCAGGACGTCCACGCAGAGAG <mark>IG</mark> TCCACCCTTGACGCCACCCTCTGCCACCCCGCAGCCCGCCACCCCCGCAGCCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCCACCCCCGCCACCCCCGCACCCCCGCCACCCCCGCCACCCCCGCACCCCCGCCACCCCCGCCACCCCCGCCACCCCCGCCACCCCCGCCACCCCCGCCACCCCCGCACCCCCGCCACCCCCGCCACCCCCGCCACCCCCGCCACCCCCGCCACCCCCGCCCCCC
Fuo gnl SRA SRR331139.224500572.3	CCAGGACCTCCACGCAGAGAGAGAGTGCCACCCTGCACCCCGCACCCCGCCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCCCGCCCCCC
REV gnl SRA SRR331142.195379237.3	CCAGGACCTCCACGCAGAGAG <mark>TG</mark> TCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331142.225926278.3 Frame 1	CCAGGACGTCCACGCAGAGAG <mark>AGAGTG</mark> TCCACCCTTGACGCCACCCTCTGCCACCCCGCAGCCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331137.151291950.3 Frame 1	CCAGGACCTCCACGCAGAGAGAGAGTGCCACCCTGCACCCCGCACCCCGCCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCACCCCCGCCCGCCCCCC
Fuo gnl SRA SRR331139.54361667.3 Frame 1	CCAGGACGTCCACGCAGAGAG <mark>AG</mark> TCCACCCTTGACGCCACCCTGTACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331141.21063563.3 G Frame 1	CCAGGACCTCCACCCCAGAGAGTGTCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331137.75678035.3 Frame 1	CCAGGACCTCCACGCAGAGAGAGAGTGTCCACCCTTGTCGCCACCCTCTACCTGCCGCCT A R T S T Q S V H P L S P A T L Y L P P
Fuo gnl SRA SRR331137.2844730.1	CCAGGACGTCCACGCAGAGAG <mark>TG</mark> TCCACCCTTGACGCCACCCTGTACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331137.80842082.1 Frame 1	CCAGGACCTCCACGCCAGAGAG <mark>TG</mark> TCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCCT A R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331142.13497177.3	NCAGGACCTCCACGCAGAGAG <mark>TG</mark> TCCACCCCTTGACGCCTGCCACCCCGTACCTGCCGCCT X R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331137.34467697.1 Frame 1	R T S T Q S V H P L T P A T L Y L P P
Fuo gnl SRA SRR331142.102265529.3 Frame 1	R T S T Q S V H P L T P A T L Y L P P
REV gnl SRA SRR331137.64477738.3 Frame 1	GOACOTCCACGCAGAGTGTCCACCOTTGACGCCTGCCACCCTCGTACCTGCCGCCT T S T Q S V H P L T P A T L Y L P P
REV gnl SRA SRR331137.26765027.1 Frame 1	ACCTCGCCGCAGAGTGTCCACCCTTGACGCCTGCCCCCCTGGACCTGCCGC T S P Q S V H P L T P A P L D L P P
FWD gnl SRA SRR331142.98410083.3 Frame 1	CTCCACGCAGAGAGAGAGTCCACCCTTGACGCCACCCTCTACCTGCCACCCCCGCCT S T Q S V H P L T P A T L Y L P P

Figure S26: SRA validation for loss of start codon in exon 1 of the *Aanat* gene from *T. manatus latirostris*.

Trichechus manatus latirostris

SRA searched: SRR331143 - Broad Institute (BI) 2011-08-17 **Sample ID:** SAMN00632092 SRR331145 - Broad Institute (BI) 2011-08-17 **Sample ID:** SAMN00632092 SRR353370 - Broad Institute (BI) 2011-10-16 **Sample ID:** SAMN00632092 SRR353371 - Broad Institute (BI) 2011-10-16 **Sample ID:** SAMN00632092

Human_ASMT - ex5	GTCCGAGGGCGAGCGGCTACAGTTCATCAAGCTCTGCAGGAGG
Frame 3	S E G E R L O F M O A L O F
Flop gi 460713528 ref NW_004444070.1 :c6504311-6437226 extraction	ICCAGGTCTGAGGACAAGCGGCTGCAGTTCATCTGAGGCCTGCAGGAGG
Frame 3	SRSEDKRRLQFMCTGCAGCTGCAGGAGG
F#0 gnl SRA SRR331145.125644409.1 Frame 3	S R S
FillD gnl SRA SRR331143.32846976.3 Frame 3	S R S E A E
REV gnl SRA SRR331143.25357114.3 Frame 3	S R S E A E Q L Q F M
REV gnl SRA SRR331145.216899669.3 Frame 3	S R S E D K R L Q F M *
Futo gnl SRA SRR353371.16345292.1	SRSEDKRAGCGGCTGCAGTTCATGTGAGGCCTGCAGGAGG
Frame 3	SRSEDKRC
Futo gnl SRA SRR353370.5295399.1	SRSEDKRAGCGGCTGCAGTTCATGTGAGGCCTGCAGGAGG
Frame 3	SRSEDKRC
Futo gnl SRA SRR353371.35464392.1 Frame 3	SRSEDKRAGCGGCTGCAGTTCATGTGAGGCCTGCAGGAGG
Futp gnl SRA SRR353371.35464301.1 Frame 3	SRSEDKRAGCGGCTGCAGTTCATGTGCGCCTGCAGGAGG
FuiD gnl SRA SRR353371.16345365.1	TCCAGGTCTGAGGACAAGCGGCTGCAGTTCATGTGAGGCCTGCAGGAGG
Frame 3	SRSEDKRRLQFM
FulD gnl SRA SRR353371.35464277.1 Frame 3	SRSEDKRAGCGGCTGCAGTTCATGTGAGCGGCTGCAGGAGG
FWD gnl SRA SRR353369.26790365.1	SRSEDKRAGCGGCTGCAGTTCATGTGAGGCCTGCAGGAGG
Frame 3	SRSEDKRC
REV gnl SRA SRR331145.27201086.3 Frame 3	S R S E D K R L Q F M G L Q E
REV gnl SRA SRR331145.111934097.3	SRSEDKRAGCGGCTGCAGTTCATGTGAGGCCTGCAGGAGG
Frame 3	SRSEDKRC
REV gnl SRA SRR331143.38330376.1	SRAEDKRAGGACGAGGGCTGCAGTTCATGTGAGGCCTGCAGGAGG
Frame 3	SRAEDKRACGCGCTGCAGGAGG
REV gnl SRA SRR331145.125644409.3	CGG TIG CAGTTCATG <mark>GGA</mark> G <mark>GC</mark> CTGCAGGAGG
Frame 3	R L Q F M <mark>G G</mark> L Q E
REV gnl SRA SRR331145.135717072.1	CGGCT <mark>G</mark> CAGTTCATG <mark>TG</mark> AG <mark>GC</mark> CTGCAGGAGG
Frame 3	R L Q F M A G L Q E
REV gnl SRA SRR331145.214434935.1	GGGCTGCAGGTCATGTGAGGCCTGCAGGAGG
Frame 3	GLQFM
REV gnl SRA SRR331143.75507579.1	GTTCATG <mark>TG</mark> AG <mark>GC</mark> CTGCAGGAGG
Frame 3	F M G L Q E

Figure S27: SRA validation for premature stop codon in exon 5 of the *Asmt* gene from *T. manatus latirostris*.

Trichechus manatus latirostris

SRA searched: SRR307134 - Broad Institute (BI) 2011-07-08 **Sample ID:** SAMN00632092 SRR307135 - Broad Institute (BI) 2011-07-08 **Sample ID:** SAMN00632092 SRR307136 - Broad Institute (BI) 2011-07-08 **Sample ID:** SAMN00632092

BTA_MTNR1A - ex2 Frame 3	LAAGATGTTCTTTGTGGATAGCTCCAATCATGTAGCAGATAGAATTAAACGCAAACCCTCTCCGTTA
FWD gi 460718260 ref NW_004444035.1 :c1630937-1619915	CAGGATGTTCTTTGTGGATAGCTCTAAAGACGTAGCAGACAGGATTAAACG-AAGCCTTCTCCGTTA
Frame 3	R M F F V D S S K D V A D R I K R-K P S P L
FWD gnl SRA SRR307135.1970962.1	CA <mark>G</mark> GATGTTCTTTGTGGATAGCTCTAA <mark>AG</mark> ACGT
Frame 3	R M F F V D S S K D
FUD gnl SRA SRR307135.87715211.1	CA <mark>G</mark> GATGTTCTTTGTGGATAGCTCTAAAGACGT
Frame 3	R M F F V D S S K D
FUD gnl SRA SRR307134.201122180.3	CAGGATGTTCTTTGTGGATAGCTCTAAAGACGTAGCAGAGGATTAAA
Frame 3	R M F F V D S S K D V A D R I K
REV gnl SRA SRR307135.121272196.1	AGGATGTTCTTTGTGGATAGCTCTAAAGACGTAGCAGAAGGATTAAACC-AAGCCTCCCGTTA
Frame 3	R M F F V D S S K D V A D R I K R- K P S P L
REV gnl SRA SRR307134.129063294.1	AGGATGTTCTTTGTGGATAGCTCTAAAGACGTAGCAGAAGGATTAAACG-AAGCCTCCCGTTA
Frame 3	R M F F V D S S K D V A D R I K 8- K P S P L
FWD gnl SRA SRR307135.89847665.3	NGNATGATNNTNNTGGNNNNCNNTAAAGACGNAGCANNCAGGATTAAACC-AAGCCTTCTCCGTTA
Frame 3	X M I/M X X X X X K D X A X R I K 8- K P S P L
REV gnl SRA SRR307135.89847665.1	AGGATGTTCTTTGTGGATAGCTCTAAAGACGTAGCAGACAGGATTAAACC-AAGCCTTCTCAGTTA
Frame 3	R M F F V D S S K D V A D R I K R-K P S Q L
REV gnl SRA SRR307136.187848432.3	TTTGTGGATAGCTC <mark>T</mark> AA <mark>AGAC</mark> GTAGCAGAACAGGATTAAACC-AAGCCTTCTCCGTTA
Frame 3	F V D S S K D V A D R I K R- K P S P L
Fuo gnl SRA SRR307135.2931995.3	AGACAGGATTAAACG-AAGCCTTCTCCGTTA
Frame 3	D R I K R- K P S P L
REV gnl SRA SRR307135.87715211.3	ACAGGAGTAAACG-AAGCCTTCTCCGTTA
Frame 3	R S K P- K P S P L

Figure S28: SRA validation for single nucleotide insertion in exon 2 of the *Mtnr1a* gene from *T*. *manatus latirostris*.

Trichechus manatus latirostris

SRA searched: SRR307134 - Broad Institute (BI) 2011-07-08 **Sample ID:** SAMN00632092 SRR307135 - Broad Institute (BI) 2011-07-08 **Sample ID:** SAMN00632092 SRR307136 - Broad Institute (BI) 2011-07-08 **Sample ID:** SAMN00632092 SRR307137 - Broad Institute (BI) 2011-07-08 **Sample ID:** SAMN00632092

Frame 1	L Y P Y P L S L V A I F H D G W	A G G G A L Q G Q R
FuD gi 460713521 ref NW_004444077.1 :c4738687-4728332 Frame 1	TGTATCCCTACCCACTACTCCTGTGGCCATCTTCCA	TGGGGGAGGCGCACTGCAAGGCCAG <mark>I</mark> G
REV gnl SRA SRR307136.128263047.1 Frame 1		
REV gnl SRA SRR307134.46044630.1 Frame 1		
REV gnl SRA SRR307134.80332115.3 Frame 1	TGTATCCCTACCCAC L Y P Y P	
REV gnl SRA SRR307137.161363281.1 Frame 1	TGTATCCCTACCCACTAATC L Y P Y P L	
REV gnl SRA SRR307137.16969891.3 Frame 1	TGTATCCTACCCACTAATC L Y P Y P L	
REV gnl SRA SRR307137.168506325.3 Frame 1		
FuD gnl SRA SRR307134.46044630.3 Frame 1		
F00 gnl SRA SRR307134.6121207.3 Frame 1	TGTATCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	
REV gnl SRA SRR307134.230015984.1 Frame 1	TGTATCCTACCCACTAATCCCTGTGGCCATCTTCCA	
REV gnl SRA SRR307136.11721591.3 Frame 1	CGTATCCCTAGCCACTAATCCCTGTGCCCATCTTCCA	
REV gnl SRA SRR307137.137954264.1 Frame 1	TTTTTCCCTACCCTCTAATCCCTGCGCCATCTGCCA F F P Y P L P V A I C 4	
F00 gnl SRA SRR307136.164052542.1 Frame 1	TGTATCCCTACCCACTAATCCCTGTGCCCATCTTCCA L Y P Y P L P V A I F H	
Fuo gnl SRA SRR307134.71788605.3 Frame 1	TGTATCCCTACCCACTAATCCCTGTGCCCCTCTTCCA	
REV gnl SRA SRR307136.97033960.1 Frame 1	TGTATCCCTACCCACTAATCCCTGTGGCCATCTTCCA	
F00 gnl SRA SRR307136.208504043.1 Frame 1	CCCMACTMATCCCTGTGCCATCTTCCA P L P V A I F H	
F00 gnl SRA SRR307136.143628956.1 Frame 1	CCCMCTMATCCCTGTGGCCATCTTCCA	
Fuo gnl SRA SRR307136.395384.1 Frame 1	CCACTAATCCCTGTGCCATCTTCCA P L P V A I F H	TGGGGGAGGCGCACTGCAAGGCCAGCG GGGGGAGGCGCACTGCAAGGCCAGGCG GGGALQGQC
FUD gnl SRA SRR307135.73828216.3 Frame 1	L X P V A I F	TGGGGGAGGCACTGCAAGGCCAG
REV gnl SRA SRR307134.6121207.1 Frame 1	P V A I F	TGGGGGAGGCGCACTGCAAGGCCAG
REV gnl SRA SRR307136.156841993.1 Frame 1	C	

Figure S29: SRA validation for frameshift mutation in exon 2 of the *Mtnr1b* gene from *T.manatus latirostris*.

Species	Accession numbers	Gene symbol	Scaffolds of H. amphibius genome	ID%	Exon number	Length	Gaps	qstart	qend	sstart	send	E-value	Bitscore
Bos taurus	NM_177509.2	AANAT	NKPW01005924.1	87.117	1	163	0	1	163	56196	56034	1.71e-44	185
Bos taurus	NM_177509.2	AANAT	NKPW01005924.1	91.720	2	157	0	163	319	55714	55558	1.68e-54	219
Bos taurus	NM_177509.2	AANAT	NKPW01005924.1	86.452	3	310	2	316	624	55168	54860	1.24e-90	339
Bos taurus	XM_024981048.1	AANAT	NKPW01005924.1	87.730	1	163	0	1	163	44541	44379	3.66e-46	191
Bos taurus	XM_024981048.1	AANAT	NKPW01005924.1	90.604	2	149	0	163	311	44059	43911	2.19e-48	198
Bos taurus	XM_024981048.1	AANAT	NKPW01005924.1	83.495	3	309	0	316	624	43518	43210	1.26e-75	289
Bos taurus	NM_001206907.2	MTNR1B	NKPW01013551.1	94.667	1	225	0	1	225	40814	40590	1.06e-93	350
Bos taurus	NM_001206907.2	MTNR1B	NKPW01013551.1	88.575	2	814	0	221	1034	26808	25995	0.0	989
Bos taurus	NM_005958.4	MTNR1A	NKPW01072562.1	84.737	1	190	6	5	185	789	978	3.79e-43	182
Bos taurus	NM_005958.4	MTNR1A	NKPW01009466.1	85.682	2	873	3	183	1053	13832	14702	0.0	917
Homo sapiens	NM_004043.2	ASMT	NKPW01039339.1	81.250	2	176	2	68	241	2180	2005	6,87E-29	141
Homo sapiens	NM_004043.2	ASMT	NKPW01075055.1	78.333	5	120	0	441	560	1215	1334	5,46E-10	78.7
Homo sapiens	NM_004043.2	ASMT	NKPW01040565.1	77.778	7	144	1	653	795	6591	6734	9,08E-13	87.9
Homo sapiens	NM_004043.2	ASMT	NKPW01040565.1	86.667	8	135	3	786	918	8397	8530	1,48E-30	147
Homo sapiens	NM 004043.2	ASMT	NKPW01040565.1	74.528	9	212	2	908	1118	2399	2609	7,02E-14	91.6

Table S1: Blast-n output of *Bos taurus* and *Homo sapiens* sequences against the *H. amphibius* genome. These values were used to find the loci of *Hippopotamus amphibius* that contains the genes *Aanat*, *Mtnr1a* and *Mtnr1b*.

Table S2: Accession numbers of the *Aanat* orthologues sequences used in phylogenetic analyses.

Species	Tree CODE	Gene ID	Accession Number Nucleotide	Gene Symbol	Lineage
Pan troglodytes	P. troglodytes	503504	NM_001012440.1	Aanat	Primate- Hominoidae
Homo sapiens	H. sapiens	15	NM_001166579.1	Aanat	Primate- Hominoidae
Pongo abelii	P. abelii	100433977	XM_002827870.4	Aanat	Primate- Hominoidae
Macaca mulatta	M. mulatta	706924	NM_001047127.1	Aanat	Primate- Cercopithecoidea
Papio anubis	P. anubis	101021111	XM_021928021.1	Aanat	Primate- Cercopithecoidea
Piliocolobus tephrosceles	P. tephrosceles	111542384	XM_023211738.1	Aanat	Primate- Cercopithecoidea
Microcebus murinus	M. murinus	105878976	XM_012778898.2	Aanat	Primate- Strepsirrhini
Bubalus bubalis	B. bubalis	102412521	XM_006045309.1	LOC102412521	Artiodactyla
Bos taurus	B. taurus	281583	NM_177509.2	Aanat	Artiodactyla
Bos taurus	B. taurus	112442710	XM_024981048.1	Aanat	Artiodactyla
Capra hircus	C. hircus	102190991	XM_018064062.1	Aanat	Artiodactyla
Capra hircus	C. hircus	100860947	NM_001285561.1	Aanat	Artiodactyla
Ovis aries	O. aries	443531	NM_001009461.1	Aanat	Artiodactyla
Odocoileus virginianus texanus	O. virginianus	110152065	XM_020915599.1	LOC110152065	Artiodactyla
Hippopotamus amphibius	H. amphibius	-	-	-	Artiodactyla
Hippopotamus amphibius	H. amphibius	-	-	-	Artiodactyla
Sus scrofa	S. scrofa	100518330	XM_005656910.3	LOC100518330	Artiodactyla
Vicugna pacos	V. pacos	102540360	XM_006199417.1	Aanat	Artiodactyla
Camelus dromedarius	C. dromedarius	105092185	XM_010983838.1	Aanat	Artiodactyla
Camelus ferus	C. ferus	102522698	XM_006176501.1	Aanat	Artiodactyla
Equus asinus	E. asinus	106830333	XM_014839904.1	Aanat	Perissodactyla- Equidea
Equus caballus	E. caballus	100058770	XM_023651967.1	Aanat	Perissodactyla- Equidea
Ceratotherium simum simum	C. simum	101404984	XM_014789296.1	LOC101404984	Perissodactyla- Rhinoceratidae

 Table S3: Accession numbers of the *Mtnr1a* orthologues sequences used in phylogenetic analyses.

Species	Tree CODE	Gene ID	Accession Number Nucleotide	Gene Symbol	Lineage
Homo sapiens	H. sapiens	4543	NM_005958.4	MTNR1A	Primate-Hominoidae
Nomascus leucogenys	N. leucogenys	100594737	XM_003271501.1	MTNR1A	Primate-Hominoidae
Pan troglodytes	P. troglodytes	471417	XM_016952653.1	MTNR1A	Primate-Hominoidae
Macaca mulatta	M. mulatta	702686	XM_001090972.3	MTNR1A	Primate- Cercopithecoidea
Papio anubis	P. anubis	101005602	XM_003899434.4	MTNR1A	Primate- Cercopithecoidea
Piliocolobus tephrosceles	P. tephrosceles	111548519	XM_023221040.2	MTNR1A	Primate- Cercopithecoidea
Microcebus murinus	M. murinus	105882472	XM_012784869.1	MTNR1A	Primate-Strepsirrhini
Bison bison bison	B. bison	104985582	XM_010835821.1	MTNR1A	Artiodactyla
Bubalus bubalis	B. bubalis	102415439	XM_006076173.1	MTNR1A	Artiodactyla
Bos taurus	B. taurus	539948	XM_002698656.4	MTNR1A	Artiodactyla
Capra hircus	C. hircus	102189094	XM_018041838.1	MTNR1A	Artiodactyla
Ovis aries	O. aries	443022	NM_001009725.1	MTNR1A	Artiodactyla
Hippopotamus amphibius	H. amphibius	-	-	MTNR1A	Artiodactyla
Equus caballus	E. caballus	100056423	XM_001490171.3	MTNR1A	Perissodactyla-Equidea
Equus przewalskii	E. przewalskii	103549958	XM_008518624.1	MTNR1A	Perissodactyla-Equidea
Equus asinus	E. asinus	106840379	XM_014855805.1	MTNR1A	Perissodactyla-Equidea
Ceratotherium simum simum	C. simum	101391490	XM_004428772.1	LOC101391490	Perissodactyla- Rhinoceratidae

Table S4: Accession numbers of the *Mtnr1b* orthologues sequences used in phylogenetic analyses.

Species	Tree CODE	Gene ID	Accession Number Nucleotide	Gene Symbol	Lineage
Pan troglodytes	P. troglodytes	466747	XM_016921786.2	Mtnr1b	Primate-Hominoidae
Homo sapiens	H. sapiens	4544	NM_005959.3	Mtnr1b	Primate-Hominoidae
Pan paniscus	P. paniscus	100973621	XM_003813777.1	Mtnr1b	Primate-Hominoidae
Papio anubis	P. anubis	100999162	XM_003910543.3	Mtnr1b	Primate- Cercopithecoidea
Piliocolobus tephrosceles	P. tephrosceles	111540700	XM_023209139.2	Mtnr1b	Primate- Cercopithecoidea
Microcebus murinus	M. murinus	105857333	XM_012739636.1	Mtnr1b	Primate-Strepsirrhini
Ovis aries	O. aries	100174786	NM_001130938.1	Mtnr1b	Artiodactyla
Capra hircus	C. hircus	102176302	XM_018042917.1	Mtnr1b	Artiodactyla
Bubalus bubalis	B. bubalis	102409257	XM_006053939.1	Mtnr1b	Artiodactyla
Bos taurus	B. taurus	528665	NM_001206907.2	Mtnr1b	Artiodactyla
Hippopotamus amphibius	H. amphibius	-	-	Mtnr1b	Artiodactyla
Camelus bactrianus	C. bactrianus	105083508	XM_010973367.1	Mtnr1b	Artiodactyla
Camelus dromedarius	C. dromedarius	105094616	XM_010986681.1	Mtnr1b	Artiodactyla
Odocoileus virginianus texanus	O. virginianus	110127734	XM_020878096.1	Mtnr1b	Artiodactyla
Equus caballus	E. caballus	100059172	XM_001917051.2	Mtnr1b	Perissodactyla- Equidea
Equus przewalskii	E. przewalskii	103548037	XM_008515596.1	Mtnr1b	Perissodactyla- Equidea
Equus asinus	E. asinus	106827482	XM_014835349.1	Mtnr1b	Perissodactyla- Equidea
Ceratotherium simum simum	C. simum	101399557	XM_004427448.1	LOC101399557	Perissodactyla- Rhinoceratidae

Gene annotations: In silico gene annotations of *Aanat*, *Mtnr1a* and *Mtnr1b* in *H. amphibius Aanat* genes of Hippopotamus amphibius annotated by Augustus Software.

----- prediction on sequence number 1 (length = 65446, name = NKPW01005924.1) -----

Gene 1

NKPW01005924.1	AUGUSTUS	gene	43218	44549	0.28	-		g2	
NKPW01005924.1	AUGUSTUS	transcr	ipt	43218	44549	0.28	-		g2.t1
NKPW01005924.1 transcript_id "	AUGUSTUS g2.t1"; gene_id	stop_co l "g2";	odon	43218	43220	•	-	0	
NKPW01005924.1 transcript_id "	AUGUSTUS g2.t1"; gene_id	termina l "g2";	al	43218	43503	0.38	-	1	
NKPW01005924.1 transcript_id "	AUGUSTUS g2.t1"; gene_id	interna l "g2";	1	43916	44066	0.37	-	2	
NKPW01005924.1 "g2.t1"; gene_id "g2"	AUGUSTUS	initial	44387	44549	0.8	-	0	transcr	ipt_id
NKPW01005924.1 "g2.t1"; gene_id "g2"	AUGUSTUS	CDS	43218	43503	0.38	-	1	transcr	ipt_id
NKPW01005924.1 "g2.t1"; gene_id "g2"	AUGUSTUS	CDS	43916	44066	0.37	-	2	transcr	ipt_id
NKPW01005924.1 "g2.t1"; gene_id "g2"	AUGUSTUS	CDS	44387	44549	0.8	-	0	transcr	ipt_id
NKPW01005924.1 transcript_id "	AUGUSTUS g2.t1"; gene_id	start_c l "g2";	odon	44547	44549		-	0	

coding sequence =

[atgtccacacagaccatccaatacatgaagcctgtggctctgcacctgccacctgggatcccagagtcgccaagccgccagcggcgccataca ctccccgccaatgagttccgctgcctcaccccagaggacgctgctgacgtgtttgagattgagcgagaggccttcatctctgtctcgggcgcctgc cccctgcaactggaccaggtccagcacttcctgaccgtgtgtcccgagctgtccctgggctgcttcgtggagggccgcctcgtgggcttcatcatt ggctccctgtgggacaaggagagactcaggcccggggggccacacagccgacctgcacttgctgggggccgcccttcggggggccgccttggg caacggttccaccctgctctggcgctctctgcaccactggatggcaggcggccgccggagggccgcccggggggcccctgggg ccttctaccagaggttcggcttccaccctgtggggcccgtggggccgccgggggcccccgtggggccccctgggggc catgcctcctgcgcaggaacagtgacagctga]

protein sequence =

[MSTQTIQYMKPVALHLPPGIPESPSRQRRHTLPANEFRCLTPEDAADVFEIEREAFISVSGACP LQLDQVQHFLTVCPELSLGCFVEGRLVAFIIGSLWDKERLRPGGHTAHLHLLAVRRPFRRLG NGSTLLWRSLHHLDGKPAVRRTSLMCEDPLVPFYQRFGFHPVGPCAVTVGSLTFTEMERSP WGHASLRRNSDS]

Gene 2

NKPW01005924.1 AUGUSTUS gene 54868 56204 0.87 - . g3

NKPW01005924.1	AUGUSTUS	transcript		54868	56204	0.87	-	•	g3.t1
NKPW01005924.1 transcript_id "	AUGUSTUS g3.t1"; gene_id	stop_co l "g3";	odon	54868	54870		-	0	
NKPW01005924.1 transcript_id "	AUGUSTUS g3.t1"; gene_id	termina l "g3";	al	54868	55173	0.93	-	0	
NKPW01005924.1 transcript_id "	AUGUSTUS g3.t1"; gene_id	interna l "g3";	1	55567	55721	0.97	-	2	
NKPW01005924.1 "g3.t1"; gene_id "g3"	AUGUSTUS ;	initial	56042	56204	0.95	-	0	transcr	ipt_id
NKPW01005924.1 "g3.t1"; gene_id "g3"	AUGUSTUS	CDS	54868	55173	0.93	-	0	transcr	ipt_id
NKPW01005924.1 "g3.t1"; gene_id "g3"	AUGUSTUS	CDS	55567	55721	0.97	-	2	transcr	ipt_id
NKPW01005924.1 "g3.t1"; gene_id "g3"	AUGUSTUS	CDS	56042	56204	0.95	-	0	transcr	ipt_id
NKPW01005924.1 transcript_id "	AUGUSTUS g3.t1"; gene_id	start_c "g3";	odon	56202	56204		-	0	

coding sequence =

protein sequence =

[MSTQTIQYMKPVALHLPPGIPESPSRQRRHTLPANEFRCLTPEDAADVFEIEREAFISVSGVCP LQLDQVQHFLTVCPELSLGCFVEGRLVAFIIGSLWDKERLTQESLTLHRPGGHTAHLHLLAV HRTFRQLGKGSTLLWRYLHHLDGQPAVRRAVLMCEDPLVPFYQRFGFHPVGLCAATVGSL TFTEMQCSLRGHASLRRNSDS]

Mtnr1a gene of Hippopotamus amphibius annotated by Augustus Software.

----- prediction on sequence number 1 (length = 65443, name = NKPW01072562.1 and NKPW01009466.1) -----

#

Constraints/Hints:

(none)

Predicted genes for sequence number 1 on both strands

start gene gl								
NKPW01072562.1_NKPW0 g1)1009466.1	AUGUSTUS	gene	770	16959	0.32	+	
NKPW01072562.1_NKPW0 . g1.t1)1009466.1	AUGUSTUS	transcr	ipt	770	16959	0.32	+
NKPW01072562.1_NKPW0 0 transcript_id)1009466.1 "g1.t1"; gene_id	AUGUSTUS "g1";	start_co	odon	770	772	•	+
NKPW01072562.1_NKPW0 transcript_id "g1.t1";)1009466.1 gene_id "g1";	AUGUSTUS	initial	770	998	1	+	0
NKPW01072562.1_NKPW0 2 transcript_id)1009466.1 "g1.t1"; gene_id	AUGUSTUS "g1";	termina	al	16091	16959	0.32	+
NKPW01072562.1_NKPW0 transcript_id "g1.t1";)1009466.1 gene_id "g1";	AUGUSTUS	CDS	770	998	1	+	0
NKPW01072562.1_NKPW0 transcript_id "g1.t1";)1009466.1 gene_id "g1";	AUGUSTUS	CDS	16091	16959	0.32	+	2
NKPW01072562.1_NKPW0 0 transcript_id)1009466.1 "g1.t1"; gene_id	AUGUSTUS "g1";	stop_co	odon	16957	16959		+

coding sequence =

protein sequence =

[MAGGPWGAPGGASKGNGSALLNASQQAPGSGEGAPSRPSWLVTTLALILIFTIVVDILGNLL VILSVYRNKKLRNAGNIFVVSLAVADLLVAVYPYPLALTSIFNNGWSLGYLHCQISGFLMGL SVIGSVFNITGIAINRYCYICHSLQYDRLYSNKNSLCYVFLIWMLTLVAIVPNLCIGTLRYDPRI YSCTFTQSISSAYTIAVVVFHFMVPMVIVIFCYLRIWVLVLQVRWRVKPDNKPKLKPQDFRN FVTMFVVFVLFAICWAPLNFIGLAVASDPASMVSRIPEWLFVASYYMAYFNSCLNAIIYGLL NQNFRQEYRRIIVSLCTAKMFFVDSSNNVADRIKCKPSPLITNHNLIKVDSV]

Mtnr1b gene of Hippopotamus amphibius annotated by Augustus Software.

----- prediction on sequence number 1 (length = 40943, name = NKPW01013551.1) -----

#

Constraints/Hints:

(none)

Predicted genes for sequence number 1 on both strands

start gene g1

NKPW(01013551.1	AUGUSTUS	gene	25975	40822	1	-		g1	
NKPW()1013551.1	AUGUSTUS	transcr	ipt	25975	40822	1	-		g1.t1
NKPW(t)1013551.1 ranscript_id "	AUGUSTUS g1.t1"; gene_id	stop_co "g1";	odon	25975	25977		-	0	
NKPW(t)1013551.1 ranscript_id "	AUGUSTUS g1.t1"; gene_id	termina "g1";	al	25975	26813	1	-	2	
NKPW("g1.t1";)1013551.1 gene_id "g1"	AUGUSTUS ;	initial	40600	40822	1	-	0	transcr	ipt_id
NKPW("g1.t1";)1013551.1 gene_id "g1";	AUGUSTUS	CDS	25975	26813	1	-	2	transcr	ipt_id
NKPW("g1.t1";)1013551.1 gene_id "g1";	AUGUSTUS ;	CDS	40600	40822	1	-	0	transcr	ipt_id
NKPW()1013551.1 ranscript_id "	AUGUSTUS g1.t1"; gene_id	start_c "g1";	odon	40820	40822		-	0	

coding sequence =

protein sequence =

[MPENRSFANCCEAGGRAESPGWTGAGGARPSGTPRPPWVAPALSAVLIVTTAVDIVGNLLV ILSVLGNRKLRNAGNLFLVSLALADLAVALYPYPLILVAIFHNGWALGEAHCKASAFVMGL SVIGSVFNITAIAINRYCYVCRSVTYHRIYRHWHTALYICLVWLLTLLAVVPNFFLGSLEYDP RIYSCTFIQTASAGYTVAVVVVHFLLPMAIVCSCYLRIWVLVLRARRKVKAESKPCLWSRNV $RSFLTMFVVFVIFAICWAPLNCIGLAVAIDPEEVAPRVPEGLFVTSYFLAYFNSCLNAIIYGLL \\ NQNFRREYKKIVSALWNPRRCQQDSSKGSQAEGPETQAPRRG]$