# 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-Marques[1•], Raquel Ruivo[1•*], Luís Q. Alves[1,2], Nelson Sousa[1], André M. Machado[1], L. Filipe C. Castro[1,2*]

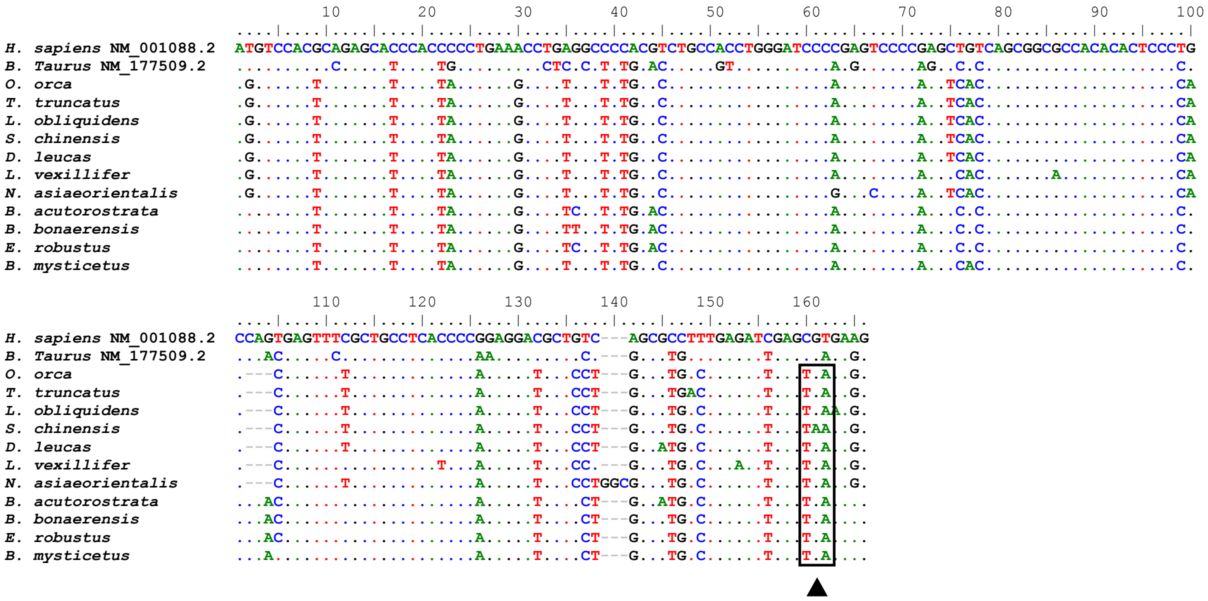
[1]CIIMAR/CIMAR - Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Portugal

[2]FCUP - 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.*



```
                                 10        20        30        40        50        60        70        80        90       100
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
H. sapiens NM_001088.2  ATGTCCACGCAGAGCACCCACCCCCTGAAACCTGAGGCCCCACGTCTGCCACCTGGGATCCCCGAGTCCCCGAGCTGTCAGCGGCGCCACACACTCCCTG
B. Taurus NM_177509.2   ..........C.....T....TG.........CTC.C.T.TG.AC.....GT..........A.G......AG..C.C....................C.
O. orca                 .G......T.......T....TA......G....T...T.TG..C.................A........A..TCAC....................CA
T. truncatus            .G......T.......T....TA......G....T...T.TG..C.................A........A..TCAC....................CA
L. obliquidens          .G......T.......T....TA......G....T...T.TG..C.................A........A..TCAC....................CA
S. chinensis            .G......T.......T....TA......G....T...T.TG..C.................A........A..TCAC....................CA
D. leucas               .G......T.......T....TA......G....T...T.TG..C.................A........A..TCAC....................CA
L. vexillifer           .G......T.......T....TA......G....T...T.TG..C.................A........A...CAC.......A............CA
N. asiaeorientalis      .G......T.......T....TA......G....T...T.TG..C.................G...C....A..TCAC....................CA
B. acutorostrata        ........T.......T....TA......G....TC..T.TG.AC.................A........A...C.C....................C.
B. bonaerensis          ........T.......T....TA......G....TT..T.TG.AC.................A........A...C.C....................C.
E. robustus             ........T.......T....TA......G....TC..T.TG.AC.................A........A...C.C....................C.
B. mysticetus           ........T.......T....TA......G....T...T.TG..C.................A........A...CAC....................C.

                               110       120       130       140       150       160
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|.
H. sapiens NM_001088.2  CCAGTGAGTTTCGCTGCCTCACCCCGGAGGACGCTGTC---AGCGCCTTTGAGATCGAGCGTGAAG
B. Taurus NM_177509.2   ...AC.....C..............AA.........C.---G...TG........T.....A..G.
O. orca                 .---C......T.............A.....T...CCT---G...TG.C......T...T.A..G.
T. truncatus            .---C......T.............A.....T...CCT---G...TGAC......T...T.A..G.
L. obliquidens          .---C......T.............A.....T...CCT---G...TG.C......T...T.AA.G.
S. chinensis            .---C......T.............A.....T...CCT---G...TG.C......T...TAA..G.
D. leucas               .---C......T.............A.....T...CCT---G..ATG.C......T...T.A..G.
L. vexillifer           .---C................T...A.....T...CC.---G...TG.C...A..T...T.A..G.
N. asiaeorientalis      .---C......T.............A.....T...CCTGGCG...TG.C......T...T.A..G.
B. acutorostrata        ...AC....................A.....T....CT---G..ATG.C......T...T.A....
B. bonaerensis          ...AC....................A.....T....CT---G...TG.C......T...T.A....
E. robustus             ...AC....................A.....T....CT---G...TG.C......T...T.A....
B. mysticetus           ...A.....................A.....T....CT---G...TG.C......T...T.A....
```

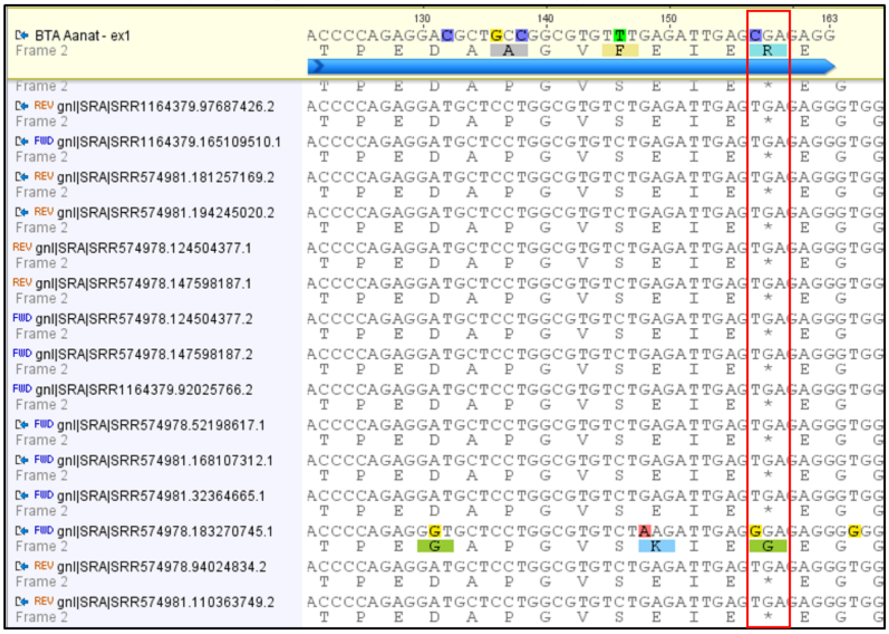
**Figure S2:** Multiple alignment of the predicted exon 1 of *Aanat* in the listed species. Conserved premature stop mutation, validated with SRA (when available), is represented in the corresponding position with a black arrow.

***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)



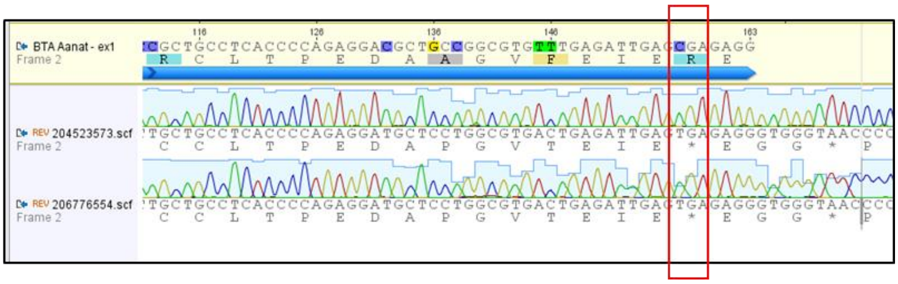
```
                                   130       140       150         163
BTA Aanat - ex1                    ACCCCAGAGGACGCTGCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 2                             T  P  E  D  A  A  G  V  F  E  I  E  R  E
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G
REV gnl|SRA|SRR1164379.97687426.2  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1164379.165109510.1 ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR574981.181257169.2  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR574981.194245020.2  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR574978.124504377.1  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G
REV gnl|SRA|SRR574978.147598187.1  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G
FWD gnl|SRA|SRR574978.124504377.2  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G
FWD gnl|SRA|SRR574978.147598187.2  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G
FWD gnl|SRA|SRR1164379.92025766.2  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G
FWD gnl|SRA|SRR574978.52198617.1   ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR574981.168107312.1  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR574981.32364665.1   ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR574978.183270745.1  ACCCCAGAGGGTGCTCCTGGCGTGTCTAAGATTGAGGGAGAGGGGGG
Frame 2                             T  P  E  G  A  P  G  V  S  K  I  E  G  E  G  G
REV gnl|SRA|SRR574978.94024834.2   ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR574981.110363749.2  ACCCCAGAGGATGCTCCTGGCGTGTCTGAGATTGAGTGAGAGGGTGG
Frame 2                             T  P  E  D  A  P  G  V  S  E  I  E  *  E  G  G
```

***Tursiops truncatus***

Trace Archive Blast

ti:1496398209

ti:1518196214



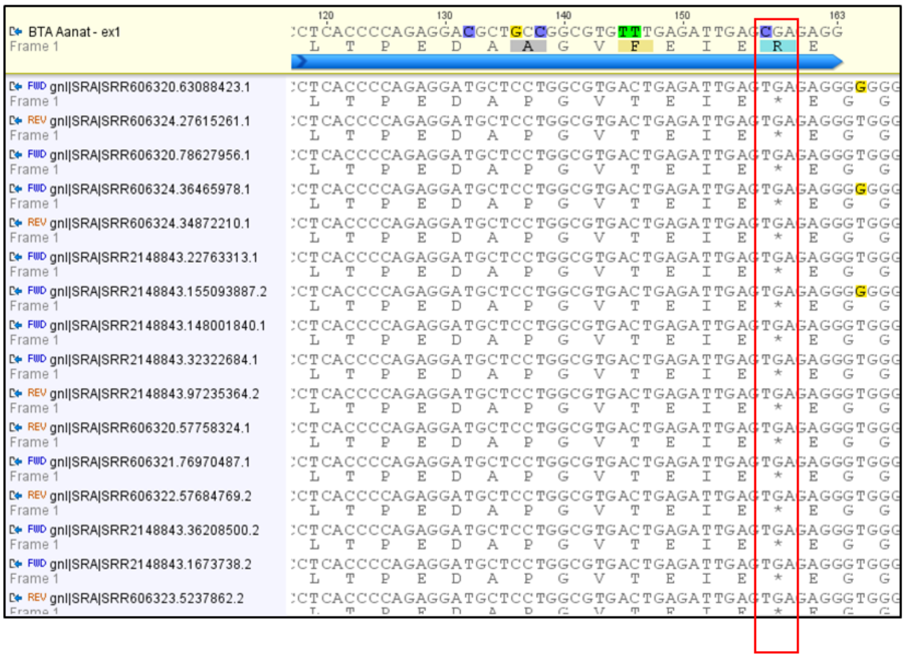
```
                    116       126       136       146         163
BTA Aanat - ex1     CGCTGCCTCACCCCAGAGGACGCTGCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 2              R  C  L  T  P  E  D  A  A  G  V  F  E  I  E  R  E
REV 204523573.scf   TGCTGCCTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGGTAACCCC
Frame 2              C  C  L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G  *  P
REV 206776554.scf   TGCTGCCTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGGTAACCCC
Frame 2              C  C  L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G  *  P
```

***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



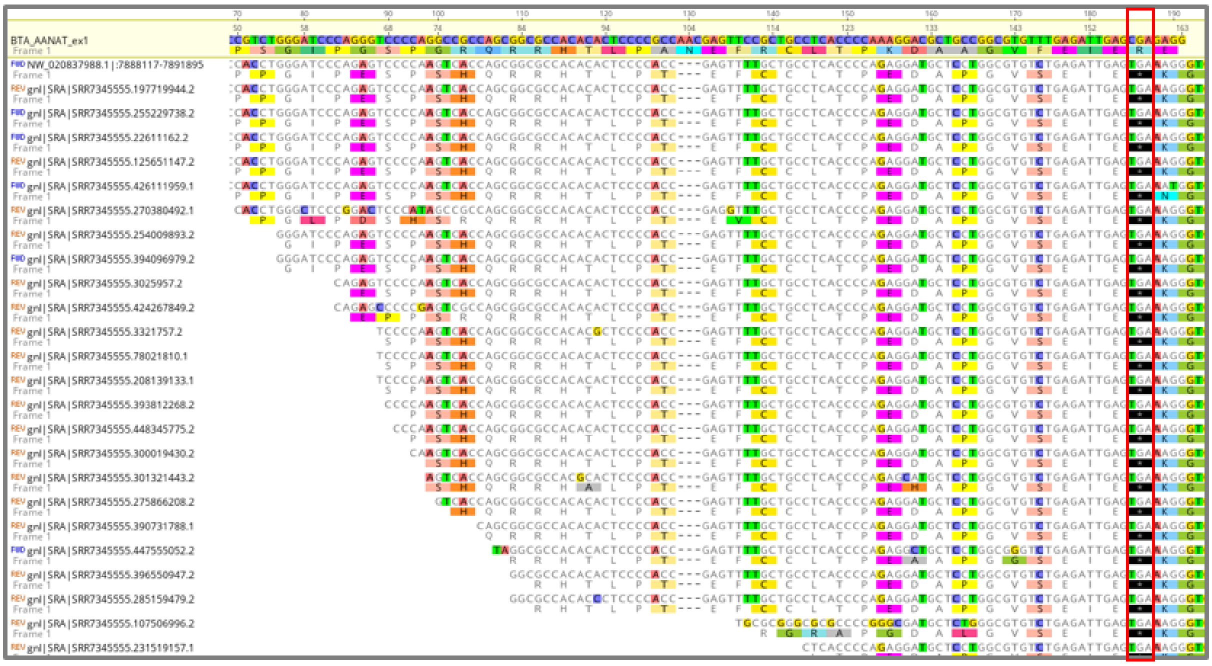
```
                                  120       130       140       150        163
BTA Aanat - ex1                   CTCACCCCAGAGGACGCTGCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 1                            L  T  P  E  D  A  A  G  V  F  E  I  E  R  E
FWD gnl|SRA|SRR606320.63088423.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGGGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
REV gnl|SRA|SRR606324.27615261.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR606320.78627956.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR606324.36465978.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGGGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
REV gnl|SRA|SRR606324.34872210.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR2148843.22763313.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR2148843.155093887.2  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGGGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR2148843.148001840.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR2148843.32322684.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
REV gnl|SRA|SRR2148843.97235364.2  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
REV gnl|SRA|SRR606320.57758324.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR606321.76970487.1  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
REV gnl|SRA|SRR606322.57684769.2  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR2148843.36208500.2  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
FWD gnl|SRA|SRR2148843.1673738.2  CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
REV gnl|SRA|SRR606323.5237862.2   CTCACCCCAGAGGATGCTCCTGGCGTGACTGAGATTGAGTGAGAGGGTGGG
Frame 1                            L  T  P  E  D  A  P  G  V  T  E  I  E  *  E  G  G
```

***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

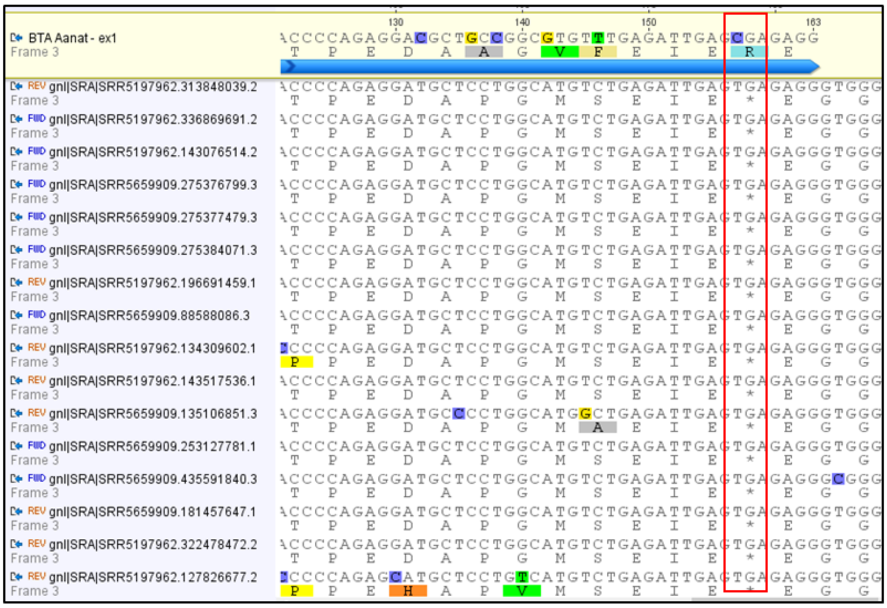


***Delphinapterus leucas***

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)



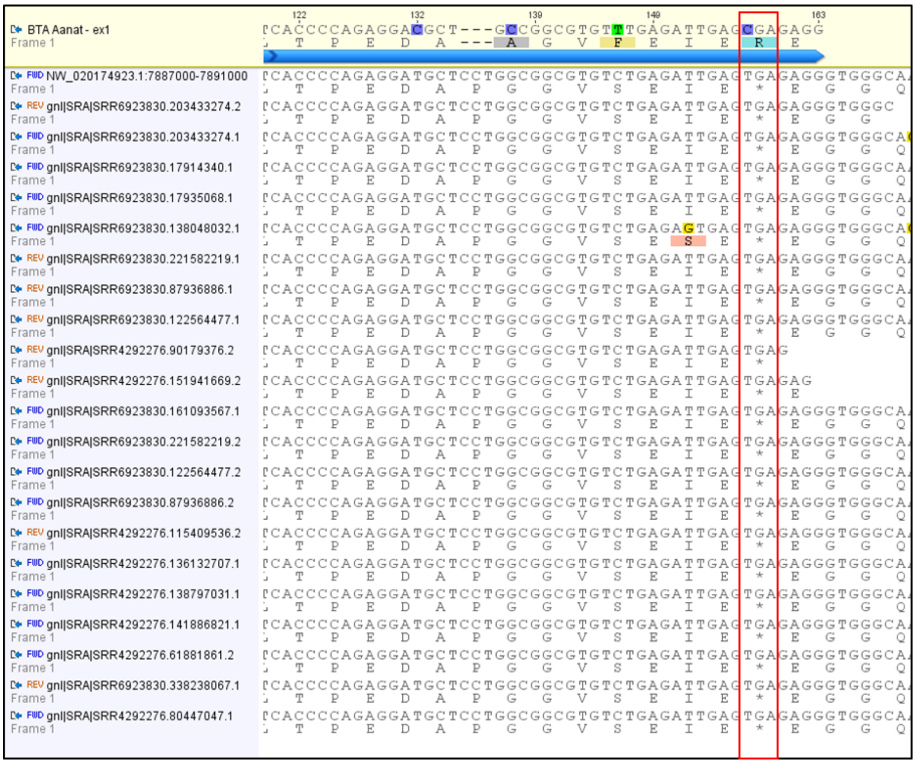
```
BTA Aanat - ex1                     CCCCAGAGGACGCTGCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 3                             T  P  E  D  A  A  G  V  F  E  I  E  R  E
REV gnl|SRA|SRR5197962.313848039.2  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5197962.336869691.2  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5197962.143076514.2  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5659909.275376799.3  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5659909.275377479.3  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5659909.275384071.3  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5197962.196691459.1  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5659909.88588086.3   CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5197962.134309602.1  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             P  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5197962.143517536.1  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5659909.135106851.3  CCCCAGAGGATGCCCCTGGCATGGCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  A  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5659909.253127781.1  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5659909.435591840.3  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGCGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5659909.181457647.1  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5197962.322478472.2  CCCCAGAGGATGCTCCTGGCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             T  P  E  D  A  P  G  M  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5197962.127826677.2  CCCCAGAGCATGCTCCTGTCATGTCTGAGATTGAGTGAGAGGGTGGG
Frame 3                             P  P  E  H  A  P  V  M  S  E  I  E  *  E  G  G
```

***Neophocaena asiaeorientalis***

SRA searched

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

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



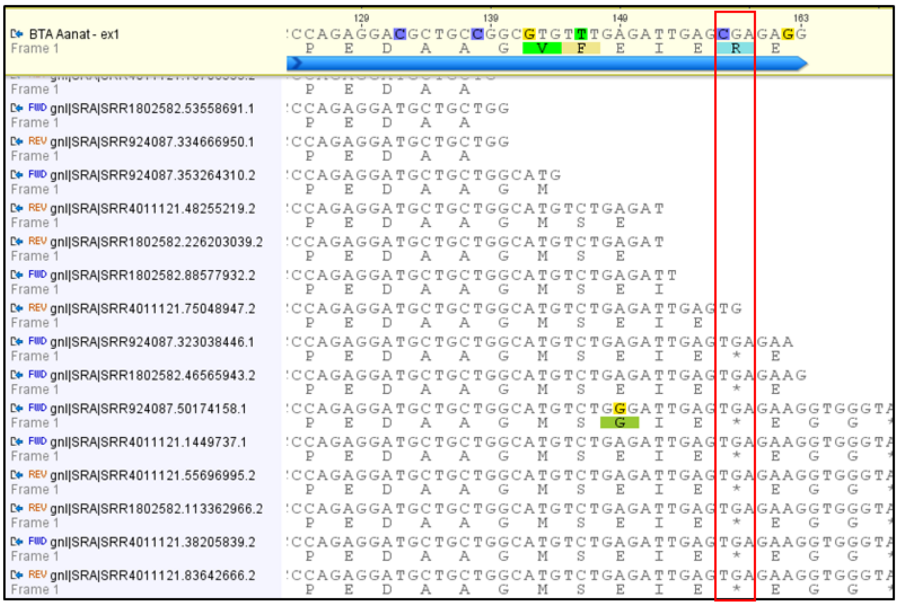
```
BTA Aanat - ex1                     CACCCCAGAGGACGCT---GCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 1                                T  P  E  D  A---  A  G  V  F  E  I  E  R  E
FWD NW_020174923.1:7887000-7891000  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
REV gnl|SRA|SRR6923830.203433274.2  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGC
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR6923830.203433274.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR6923830.17914340.1   CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR6923830.17935068.1   CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR6923830.138048032.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGAGTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  S  E  *  E  G  G  Q
REV gnl|SRA|SRR6923830.221582219.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
REV gnl|SRA|SRR6923830.87936886.1   CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
REV gnl|SRA|SRR6923830.122564477.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
REV gnl|SRA|SRR4292276.90179376.2   CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAG
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *
REV gnl|SRA|SRR4292276.151941669.2  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAG
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E
FWD gnl|SRA|SRR6923830.161093567.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR6923830.221582219.2  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR6923830.122564477.2  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR6923830.87936886.2   CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
REV gnl|SRA|SRR4292276.115409536.2  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR4292276.136132707.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR4292276.138797031.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR4292276.141886821.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR4292276.61881861.2   CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
REV gnl|SRA|SRR6923830.338238067.1  CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
FWD gnl|SRA|SRR4292276.80447047.1   CACCCCAGAGGATGCTCCTGGCGGCGTGTCTGAGATTGAGTGAGAGGGTGGGCA
Frame 1                                T  P  E  D  A  P  G  G  V  S  E  I  E  *  E  G  G  Q
```

***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)

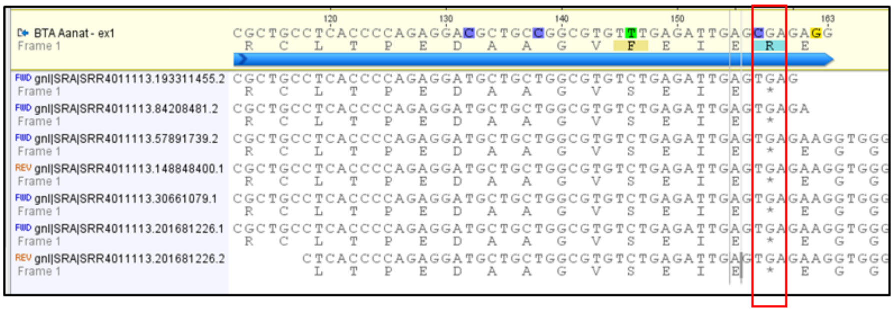


```
                                        129       139       149           163
BTA Aanat - ex1                       :CCAGAGGACGCTGCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 1                                 P  E  D  A  A  G  V  F  E  I  E  R  E
Frame 1                                 P  E  D  A  A
FWD gnl|SRA|SRR1802582.53558691.1     :CCAGAGGATGCTGCTGG
Frame 1                                 P  E  D  A  A
REV gnl|SRA|SRR924087.334666950.1     :CCAGAGGATGCTGCTGG
Frame 1                                 P  E  D  A  A
FWD gnl|SRA|SRR924087.353264310.2     :CCAGAGGATGCTGCTGGCATG
Frame 1                                 P  E  D  A  A  G  M
REV gnl|SRA|SRR4011121.48255219.2     :CCAGAGGATGCTGCTGGCATGTCTGAGAT
Frame 1                                 P  E  D  A  A  G  M  S  E
REV gnl|SRA|SRR1802582.226203039.2    :CCAGAGGATGCTGCTGGCATGTCTGAGAT
Frame 1                                 P  E  D  A  A  G  M  S  E
FWD gnl|SRA|SRR1802582.88577932.2     :CCAGAGGATGCTGCTGGCATGTCTGAGATT
Frame 1                                 P  E  D  A  A  G  M  S  E  I
REV gnl|SRA|SRR4011121.75048947.2     :CCAGAGGATGCTGCTGGCATGTCTGAGATTGAGTG
Frame 1                                 P  E  D  A  A  G  M  S  E  I  E
FWD gnl|SRA|SRR924087.323038446.1     :CCAGAGGATGCTGCTGGCATGTCTGAGATTGAGTGAGAA
Frame 1                                 P  E  D  A  A  G  M  S  E  I  E  *  E
FWD gnl|SRA|SRR1802582.46565943.2     :CCAGAGGATGCTGCTGGCATGTCTGAGATTGAGTGAGAAG
Frame 1                                 P  E  D  A  A  G  M  S  E  I  E  *  E
FWD gnl|SRA|SRR924087.50174158.1      :CCAGAGGATGCTGCTGGCATGTCTGGGATTGAGTGAGAAGGTGGGTA
Frame 1                                 P  E  D  A  A  G  M  S  G  I  E  *  E  G  G  *
FWD gnl|SRA|SRR4011121.1449737.1      :CCAGAGGATGCTGCTGGCATGTCTGAGATTGAGTGAGAAGGTGGGTA
Frame 1                                 P  E  D  A  A  G  M  S  E  I  E  *  E  G  G  *
REV gnl|SRA|SRR4011121.55696995.2     :CCAGAGGATGCTGCTGGCATGTCTGAGATTGAGTGAGAAGGTGGGTA
Frame 1                                 P  E  D  A  A  G  M  S  E  I  E  *  E  G  G  *
REV gnl|SRA|SRR1802582.113362966.2    :CCAGAGGATGCTGCTGGCATGTCTGAGATTGAGTGAGAAGGTGGGTA
Frame 1                                 P  E  D  A  A  G  M  S  E  I  E  *  E  G  G  *
FWD gnl|SRA|SRR4011121.38205839.2     :CCAGAGGATGCTGCTGGCATGTCTGAGATTGAGTGAGAAGGTGGGTA
Frame 1                                 P  E  D  A  A  G  M  S  E  I  E  *  E  G  G  *
REV gnl|SRA|SRR4011121.83642666.2     :CCAGAGGATGCTGCTGGCATGTCTGAGATTGAGTGAGAAGGTGGGTA
Frame 1                                 P  E  D  A  A  G  M  S  E  I  E  *  E  G  G  *
```

***Balaenoptera bonaerensis***

SRA searched

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



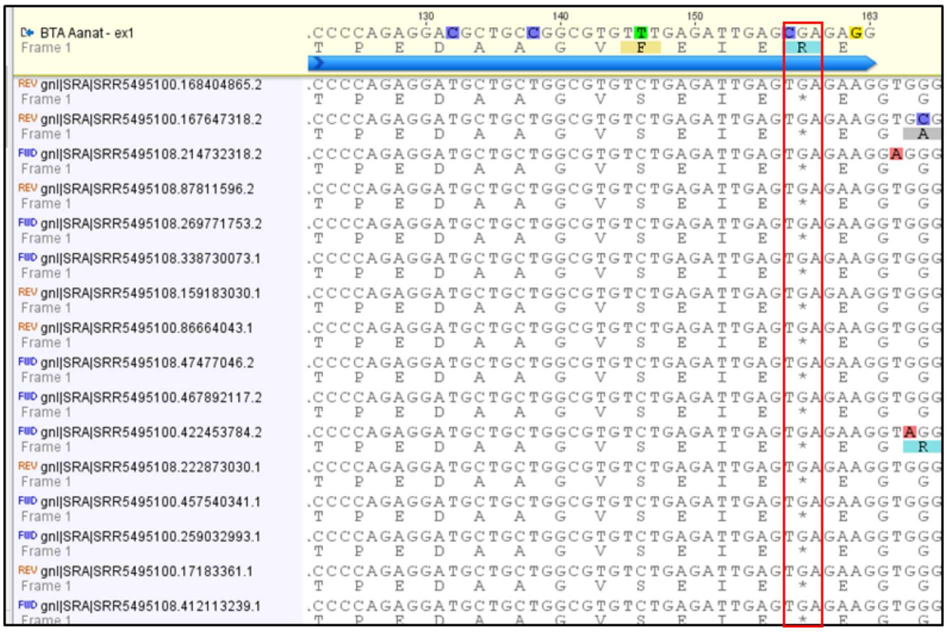
```
                                            120       130       140       150          163
BTA Aanat - ex1                     CGCTGCCTCACCCCAGAGGACGCTGCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 1                              R  C  L  T  P  E  D  A  A  G  V  F  E  I  E  R  E
FWD gnl|SRA|SRR4011113.193311455.2  CGCTGCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAG
Frame 1                              R  C  L  T  P  E  D  A  A  G  V  S  E  I  E  *
FWD gnl|SRA|SRR4011113.84208481.2   CGCTGCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGA
Frame 1                              R  C  L  T  P  E  D  A  A  G  V  S  E  I  E  *
FWD gnl|SRA|SRR4011113.57891739.2   CGCTGCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                              R  C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR4011113.148848400.1  CGCTGCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                              R  C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR4011113.30661079.1   CGCTGCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                              R  C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR4011113.201681226.1  CGCTGCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                              R  C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR4011113.201681226.2        CTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                                    L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
```

***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)



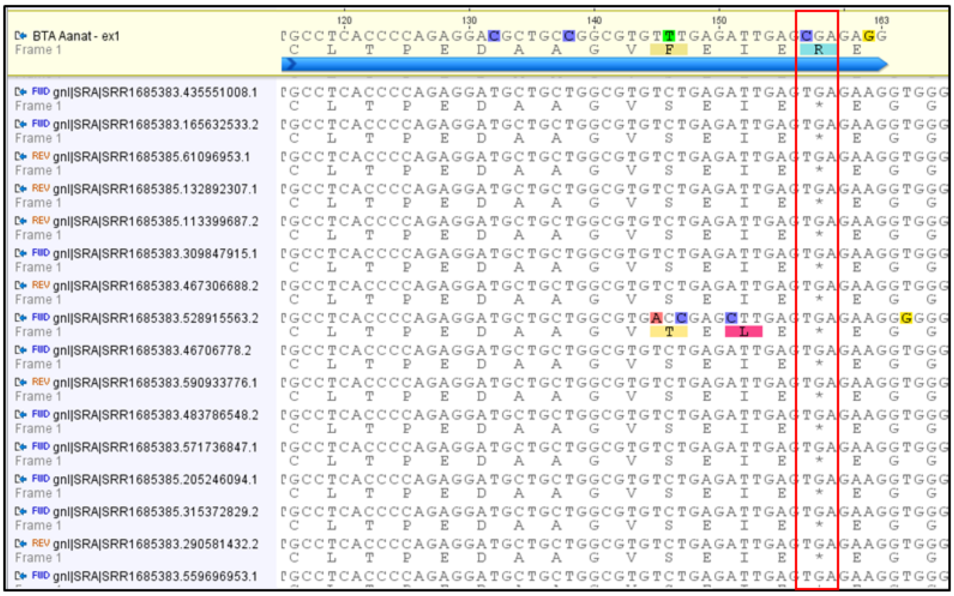
```
                                  130          140          150          163
BTA Aanat - ex1                 .CCCCAGAGGACGCTGCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 1                          T  P  E  D  A  A  G  V  F  E  I  E  R  E
REV gnl|SRA|SRR5495100.168404865.2  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5495100.167647318.2  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGCG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  A
FWD gnl|SRA|SRR5495108.214732318.2  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGAGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5495108.87811596.2   .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5495108.269771753.2  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5495108.338730073.1  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5495108.159183030.1  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5495100.86664043.1   .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5495108.47477046.2   .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5495100.467892117.2  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5495100.422453784.2  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTAGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  R
REV gnl|SRA|SRR5495108.222873030.1  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5495100.457540341.1  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5495100.259032993.1  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR5495100.17183361.1   .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR5495108.412113239.1  .CCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                          T  P  E  D  A  A  G  V  S  E  I  E  *  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)



```
                               120          130          140          150          163
BTA Aanat - ex1                GCCTCACCCCAGAGGACGCTGCCGGCGTGTTTGAGATTGAGCGAGAGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  F  E  I  E  R  E
FWD gnl|SRA|SRR1685383.435551008.1  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1685383.165632533.2  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR1685385.61096953.1   GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR1685385.132892307.1  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR1685385.113399687.2  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1685383.309847915.1  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR1685383.467306688.2  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1685383.528915563.2  GCCTCACCCCAGAGGATGCTGCTGGCGTGACCGAGCTTGAGTGAGAAGGGGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  T  E  L  E  *  E  G  G
FWD gnl|SRA|SRR1685383.46706778.2   GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR1685383.590933776.1  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1685383.483786548.2  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1685383.571736847.1  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1685385.205246094.1  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1685385.315372829.2  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
REV gnl|SRA|SRR1685383.290581432.2  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
Frame 1                         C  L  T  P  E  D  A  A  G  V  S  E  I  E  *  E  G  G
FWD gnl|SRA|SRR1685383.559696953.1  GCCTCACCCCAGAGGATGCTGCTGGCGTGTCTGAGATTGAGTGAGAAGGTGGG
```

**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.*

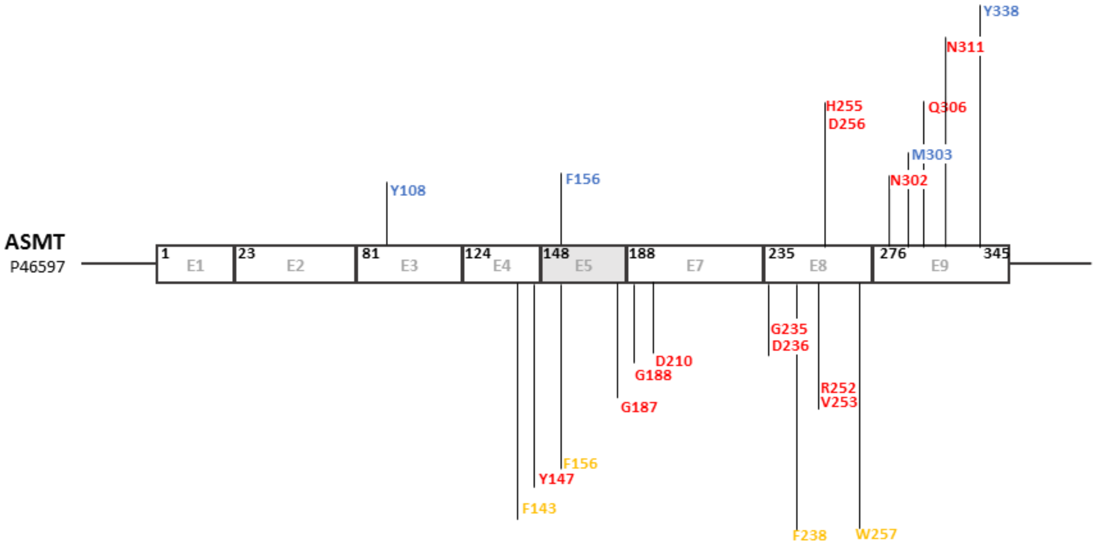
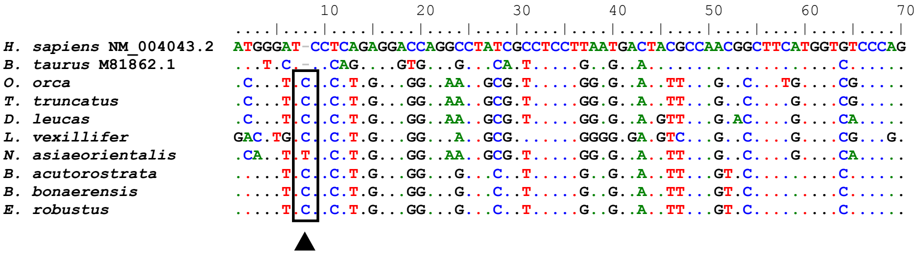


**Figure S5:** Schematic representation of the Human ASMT isoform 1 (not to scale). On top, in blue, relative location of conserved residues in proximity to the hydroxyl side of *N*-acetyl serotonin (NAS), in red residues that establish H-bonds to NAS. On the bottom in orange conserved aromatic residues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



```
                                  10        20        30        40        50        60        70
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
H. sapiens NM_004043.2    ATGGGAT-CCTCAGAGGACCAGGCCTATCGCCTCCTTAATGACTACGCCAACGGCTTCATGGTGTCCCAG
B. taurus M81862.1        ...T.C.-..CAG....GTG...G...CA.T.....G..G..A....................C......
O. orca                   .C...T.C..C.T.G...GG..AA..GCG.T.....GG.G..A..TT...G..C...TG....CG.....
T. truncatus              .C...T.C..C.T.G...GG..AA..GCG.T.....GG.G..A..TT...G..C....G....CG.....
D. leucas                 .C...T.C..C.T.G...GG..AA..GCG.T.....GG.G..A.GTT...G.AC....G....CA.....
L. vexillifer             GAC.TG.C..C.T.G...GG...A..GCG.......GGGG.GA.GTC...G..C....G....CG...G.
N. asiaeorientalis        .CA..T.T..C.T.G...GG..AA..GCG.T.....GG.G..A..TT...G..C....G....CA.....
B. acutorostrata          .....T.C..C.T.G...GG...G...C..T.....G..G..A..TT...GT.C.........C......
B. bonaerensis            .....T.C..C.T.G...GG...G...C..T.....G..G..A..TT...GT.C.........C......
E. robustus               .....T.C..C.T.G...GG...G...C..T.....G..G..A..TT...GT.C.........C......
```

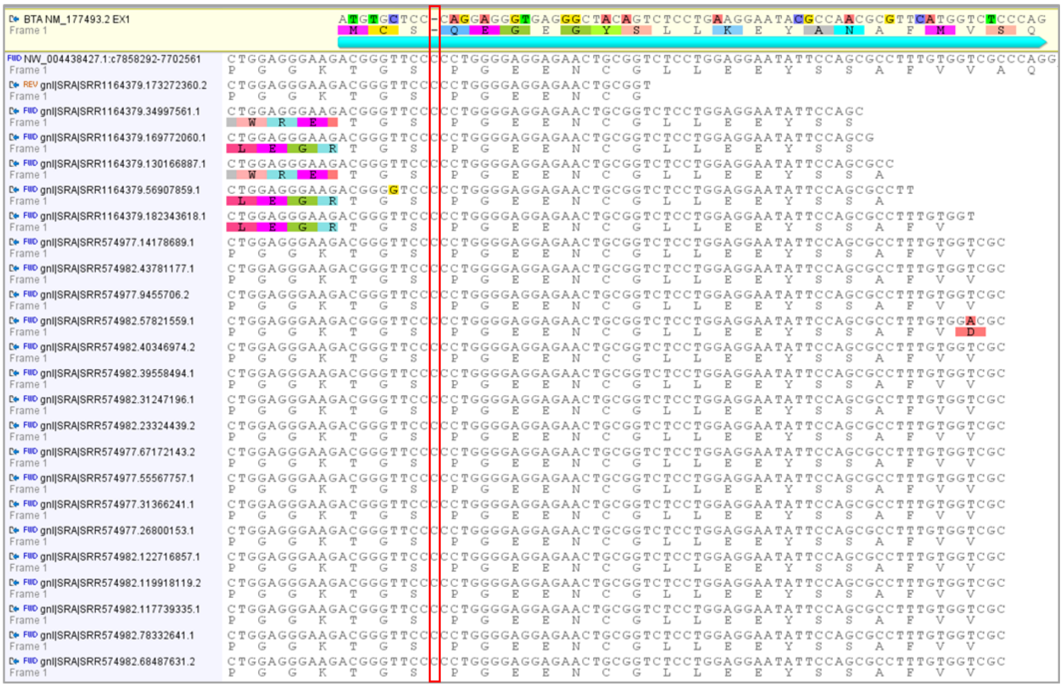
**Figure S6:** Multiple alignment of the predicted exon 1 of *Asmt* in the listed species. Conserved single nucleotide insertion, validated with SRA (when available), is represented in the corresponding position with a black arrow.

***Orcinus orca***

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)



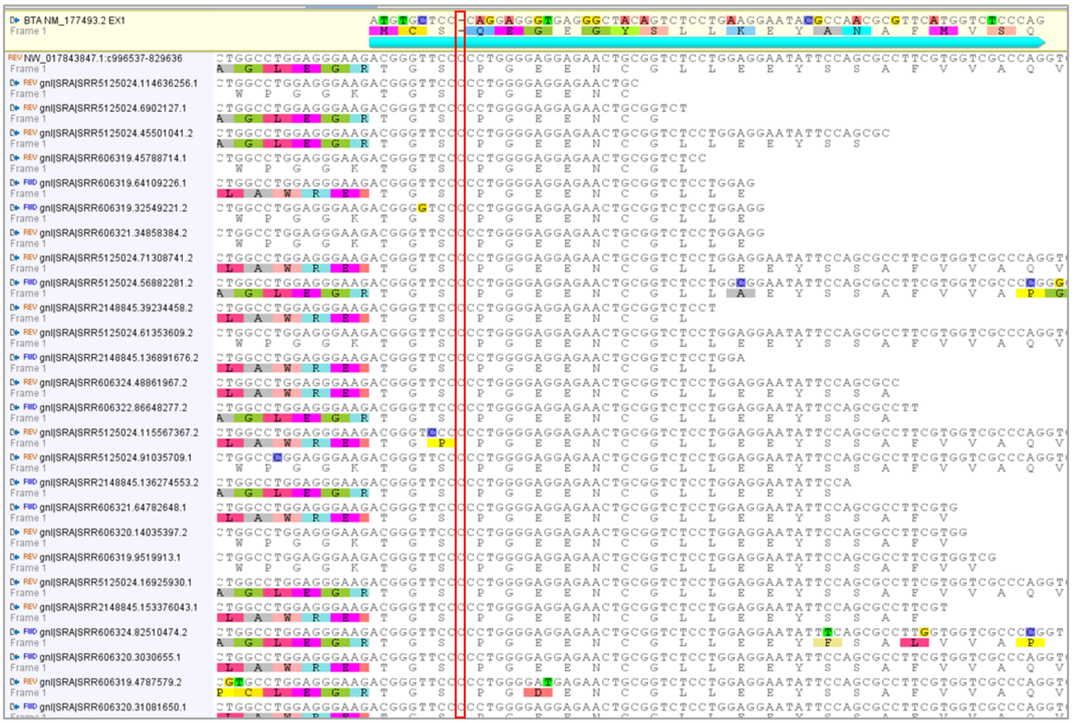
***Tursiops truncatus***

SRA searched

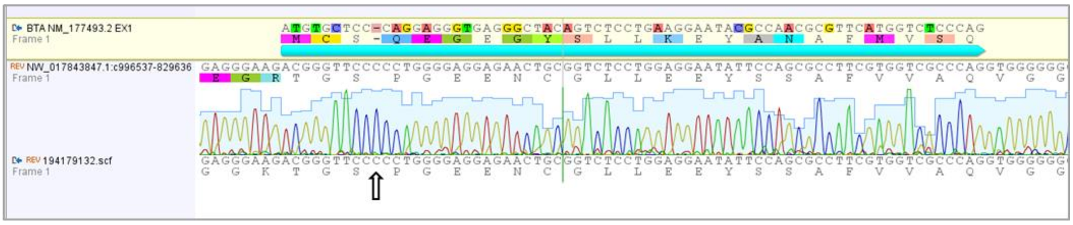
**SRR5125024 -** National Institute of Standards and Technology 2016-12-27 **Sample ID:**SAMN06114300

**SRR606319/20/21/22/24**- Baylor College of Medicine 2015-07-22 **Sample ID:** SAMN00000070

**SRR2148845 -** Beijing Genome Institute 2016-08-10 **Sample ID**:SAMN03968479

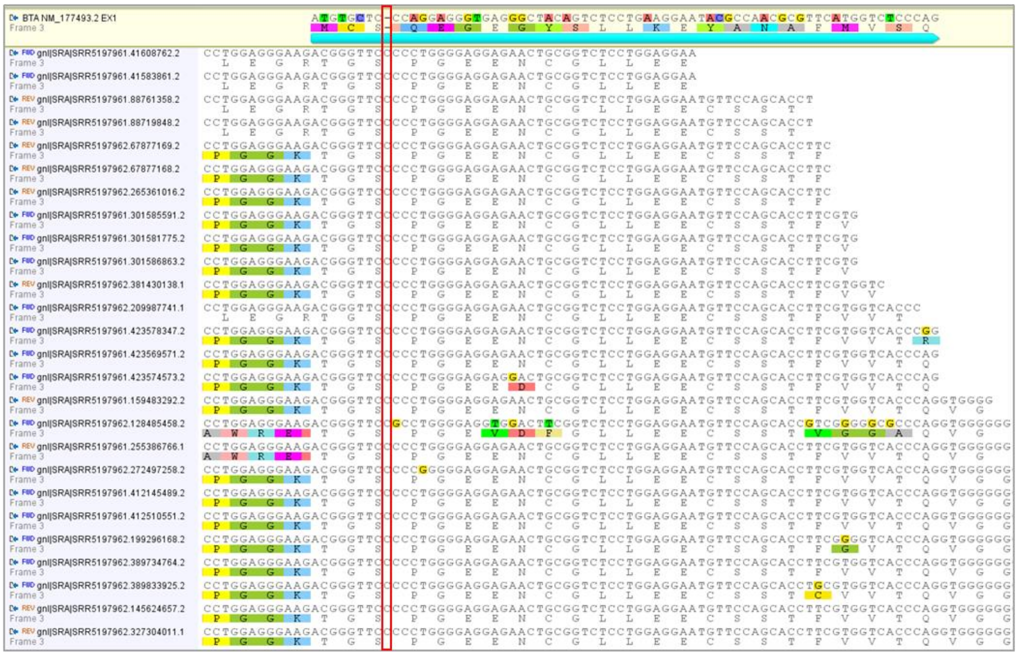


***Tursiops truncatus***
Trace Archive Blast
ti:1416172502



```
BTA NM_177493.2 EX1                    ATGTGCTCC-CAGGAGGGTGAGGGCTACAGTCTCCTGAAGGAATACGCCAACGCGTTCATGGTCTCCCAG
Frame 1                                 M  C  S  - Q  E  G  E  G  Y  S  L  L  K  E  Y  A  N  A  F  M  V  S  Q
REV NW_017843847.1:c996537-829636  GAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATATTCCAGCGCCTTCGTGGTCGCCCAGGTGGGGGG
Frame 1                             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  G  G
REV 194179132.scf                  GAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATATTCCAGCGCCTTCGTGGTCGCCCAGGTGGGGGG
Frame 1                             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  Q  V  G  G
```

***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)



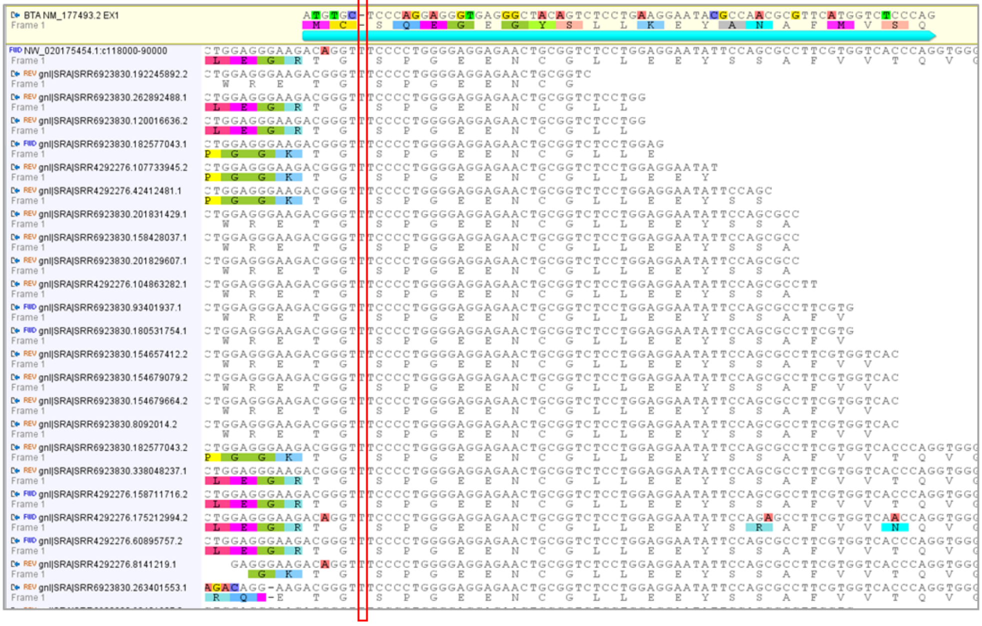
```
BTA NM_177493.2 EX1                               ATGTGCTC-CCAGGAGGGTGAGGGCTACAGTCTCCTGAAGGAATACGCCAACGCGTTCATGGTCTCCCAG
Frame 3                                            M  C  S - Q  E  G  E  G  Y  S  L  L  K  E  Y  A  N  A  F  M  V  S  Q
FWD gnl|SRA|SRR5197961.41608762.2    CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAA
Frame 3                                L  E  G  R  T  G  S  P  G  E  E  N  C  G  L  L  E  E
FWD gnl|SRA|SRR5197961.41583861.2    CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAA
Frame 3                                L  E  G  R  T  G  S  P  G  E  E  N  C  G  L  L  E  E
REV gnl|SRA|SRR5197961.88761358.2    CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCT
Frame 3                                L  E  G  R  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T
REV gnl|SRA|SRR5197961.88719848.2    CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCT
Frame 3                                L  E  G  R  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T
REV gnl|SRA|SRR5197962.67877169.2    CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTC
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F
REV gnl|SRA|SRR5197962.67877168.2    CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTC
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F
REV gnl|SRA|SRR5197962.265361016.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTC
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F
FWD gnl|SRA|SRR5197961.301585591.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V
FWD gnl|SRA|SRR5197961.301581775.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V
FWD gnl|SRA|SRR5197961.301586863.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V
REV gnl|SRA|SRR5197962.381430138.1   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTC
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V
FWD gnl|SRA|SRR5197962.209987741.1   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCC
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
FWD gnl|SRA|SRR5197961.423578347.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  R
FWD gnl|SRA|SRR5197961.423569571.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q
FWD gnl|SRA|SRR5197961.423574573.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGGACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  D  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q
REV gnl|SRA|SRR5197961.159483292.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q  V  G
FWD gnl|SRA|SRR5197962.128485458.2   CCTGGAGGGAAGACGGGTTCCGCCTGGGGAGGTGGACTTCGGTCTCCTGGAGGAATGTTCCAGCACCGTCGGGGGCGCCCAGGTGGGGGG
Frame 3                               A  W  R  E  T  G  S  P  G  E  V  D  F  G  L  L  E  E  C  S  S  T  V  G  G  A  Q  V  G
REV gnl|SRA|SRR5197961.255386766.1   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3                               A  W  R  E  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q  V  G
FWD gnl|SRA|SRR5197962.272497258.2   CCTGGAGGGAAGACGGGTTCCCCCGGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q  V  G  G
FWD gnl|SRA|SRR5197961.412145489.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q  V  G  G
FWD gnl|SRA|SRR5197961.412510551.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q  V  G  G
FWD gnl|SRA|SRR5197962.199296168.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGGGGTCACCCAGGTGGGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  G  V  T  Q  V  G  G
FWD gnl|SRA|SRR5197962.389734764.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q  V  G  G
FWD gnl|SRA|SRR5197962.389833925.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTGCGTGGTCACCCAGGTGGGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  C  V  V  T  Q  V  G  G
REV gnl|SRA|SRR5197962.145624657.2   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q  V  G  G
REV gnl|SRA|SRR5197962.327304011.1   CCTGGAGGGAAGACGGGTTCCCCCTGGGGAGGAGAACTGCGGTCTCCTGGAGGAATGTTCCAGCACCTTCGTGGTCACCCAGGTGGGGGG
Frame 3                               P  G  G  K  T  G  S  P  G  E  E  N  C  G  L  L  E  E  C  S  S  T  F  V  V  T  Q  V  G  G
```

***Neophocaena asiaeorientalis***

SRA searched

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

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

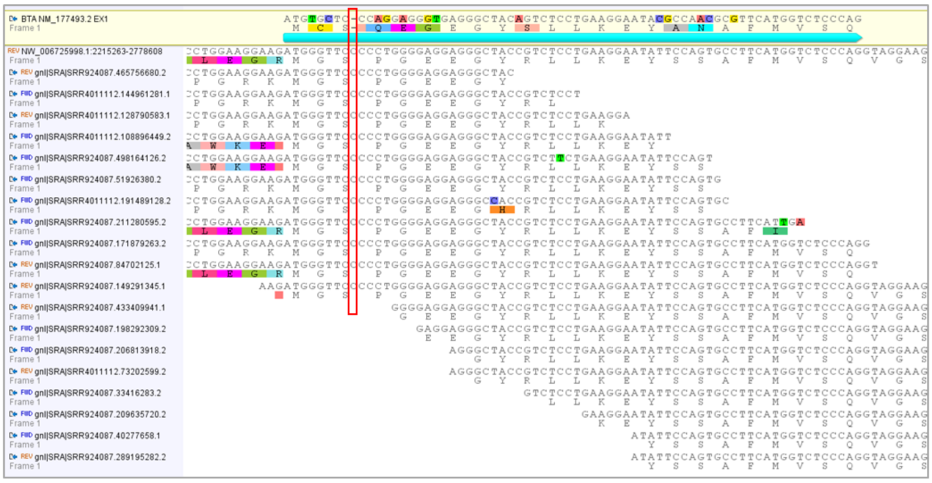


***Balaenoptera acutorostrata***

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)



***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)

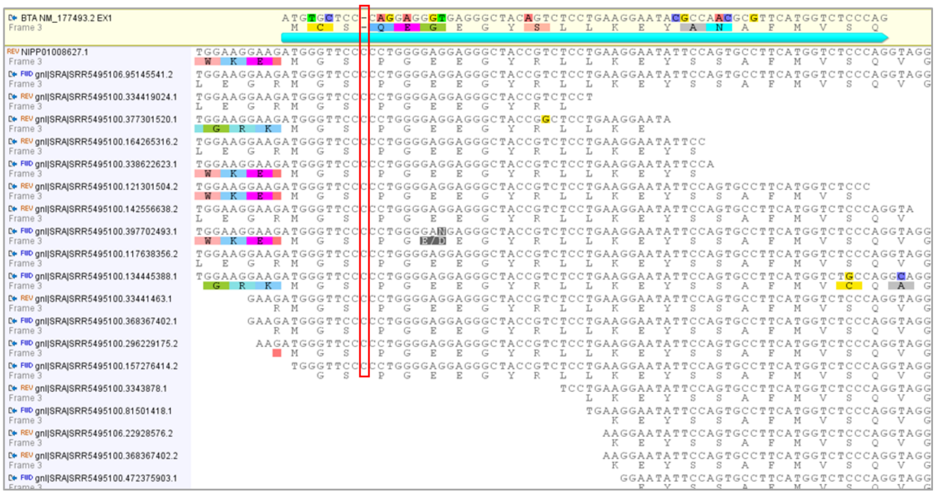


**Figure S7:** SRA validation for conserved frameshift mutation (red box) in exon 1 of the *Asmt* gene from *O. orca*, *T. truncatus*, *D. leucas*, *N. asiaeorientalis*, *B. acutorostrata* and *E. robustus*.

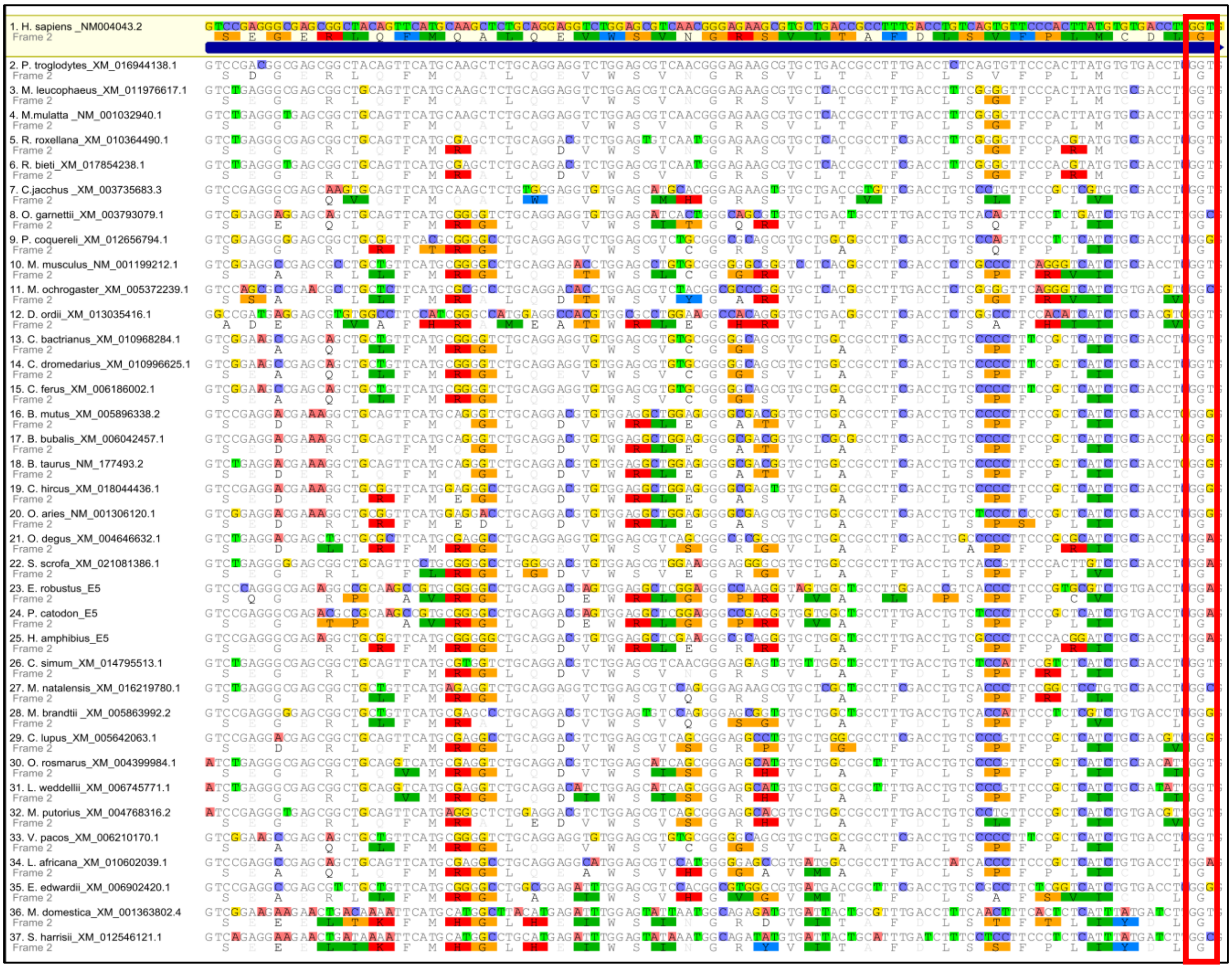


**Figure S8:** Alignment of exon 5 of *Asmt* in multiple mammals, red box indicates conserved Gly187 participating in the active site of ASMT.

***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

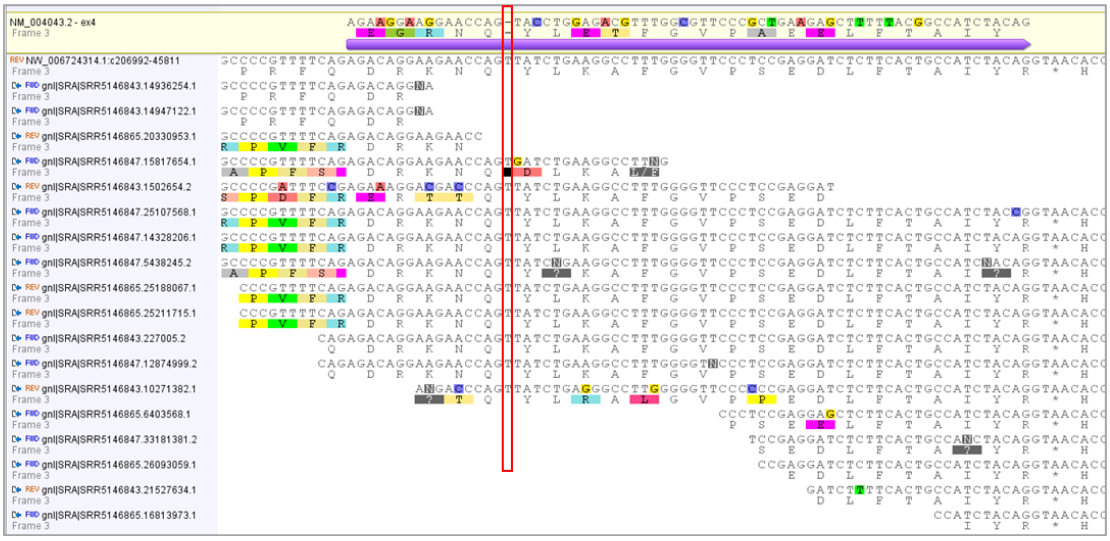


**Figure S9:** SRA validation for frameshift mutation (red box) in exon 4 of *P. catodon Asmt*.

***Lagenorhynchus obliquidens***

129

SRA searched

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

130

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

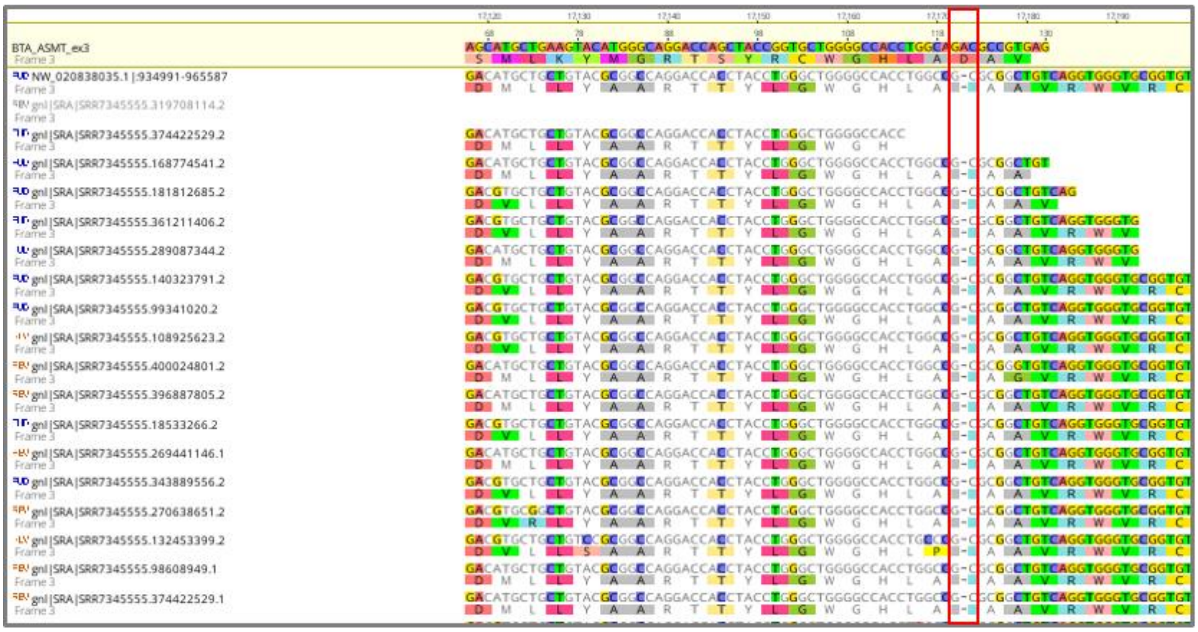


**Figure S10:** SRA validation for frameshift mutation (red box) in exon 3 of *L. obliquidens Asmt*.

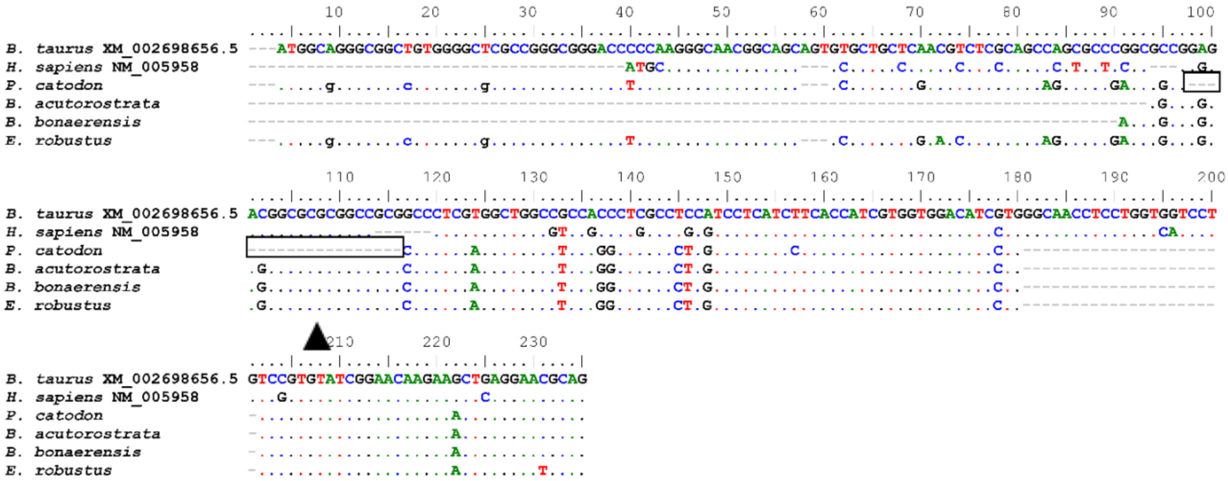


**Figure S11:** Multiple alignment of the predicted *Mtnr1a* relic sequences (exon 1) in the listed species. Nucleotide insertion of *P. catodon*, validated with SRA, is represented in the corresponding position with a black arrow

***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

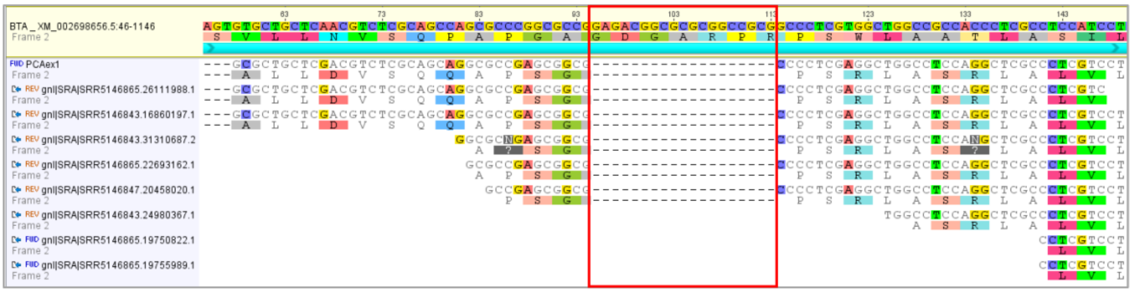


**Figure S12:** SRA validation for 19 nucleotide deletion in exon 1 of *P. catodon Mtnr1a*.

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

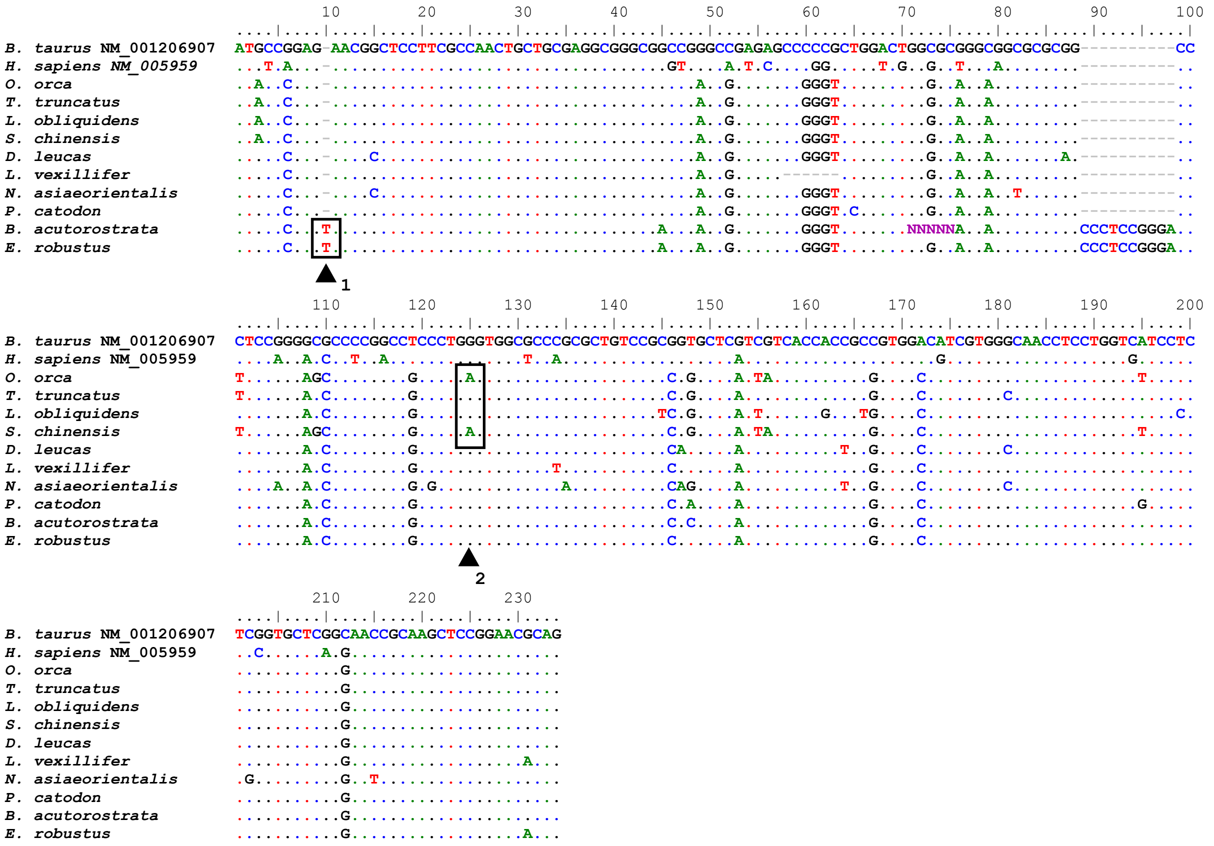


**Figure S14:** Multiple alignment of the predicted exon 1 of *Mtnr1b* in the listed species. Nucleotide insertion, generating a premature stop codon, in *B. acutorostrata* and *E. robustus* and premature stop codon retrieved in *O. orca* and *S. chinensis*, mutations validated with SRA (when available),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

***Orcinus orca*** Stop codon exon 1

SRA searched

**SRR1164379**- University of Durham 2014-02-13 **Sample ID:** SAMN02595096 (AR-Genome)

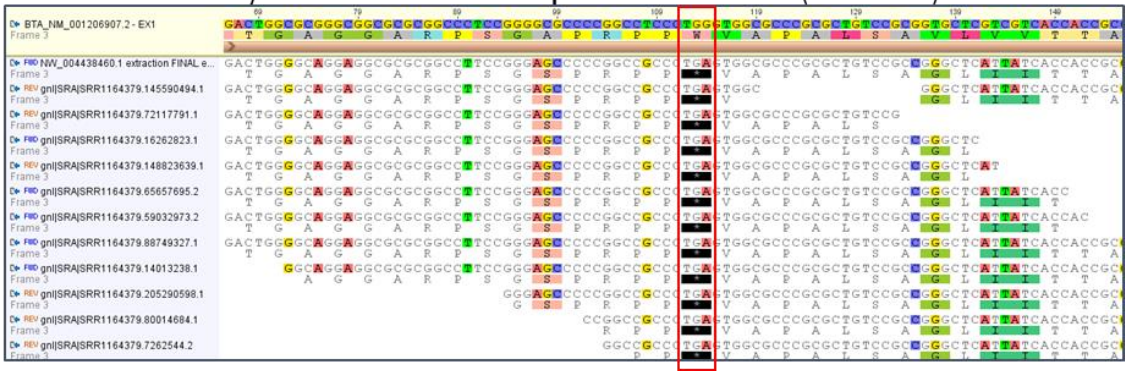
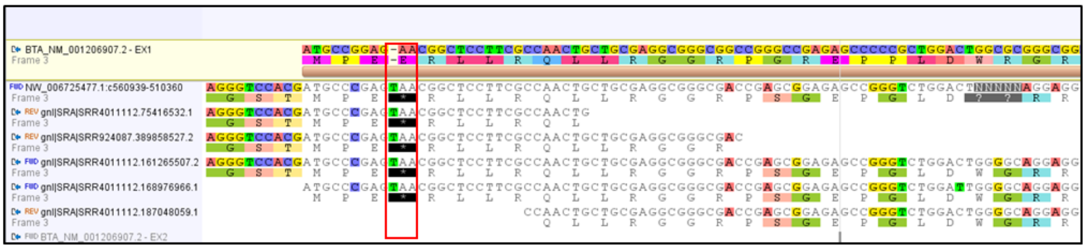


**Figure S15:** SRA validation for stop mutation (red box) in exon 1 of the *Mntr1b* gene from *O. orca*.

***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)



***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)

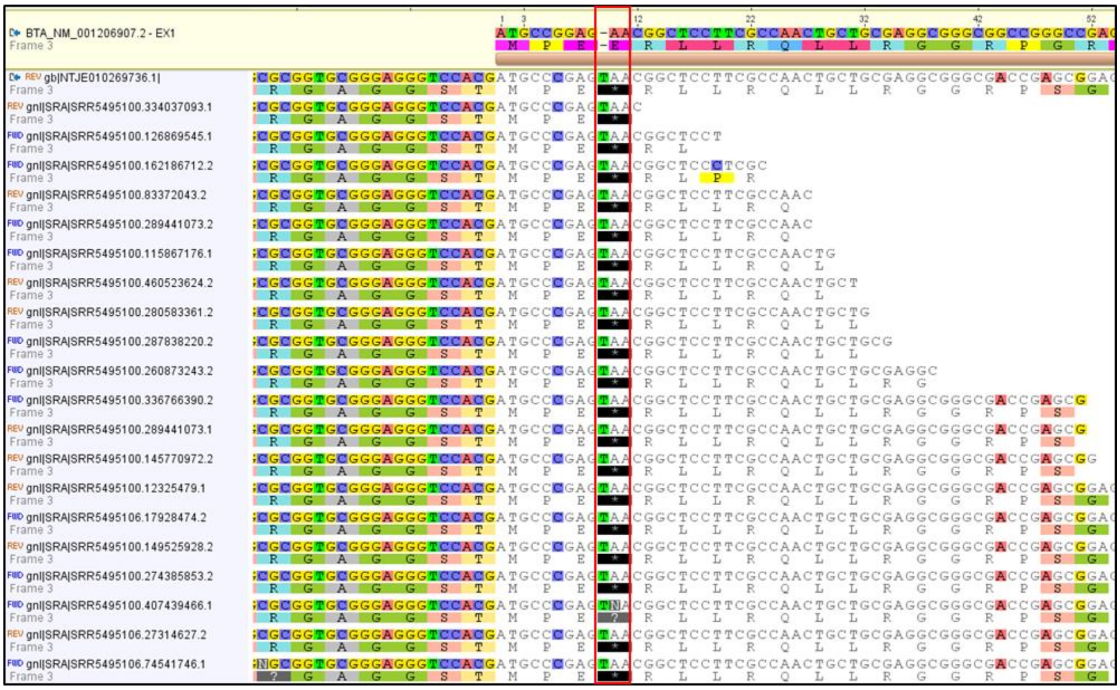


**Figure S16:** SRA validation for nucleotide insertion generating a premature stop mutation (red box) in exon 1 of the *Mntr1b* gene from *B. acutorostrata* and *E. robustus*.

***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

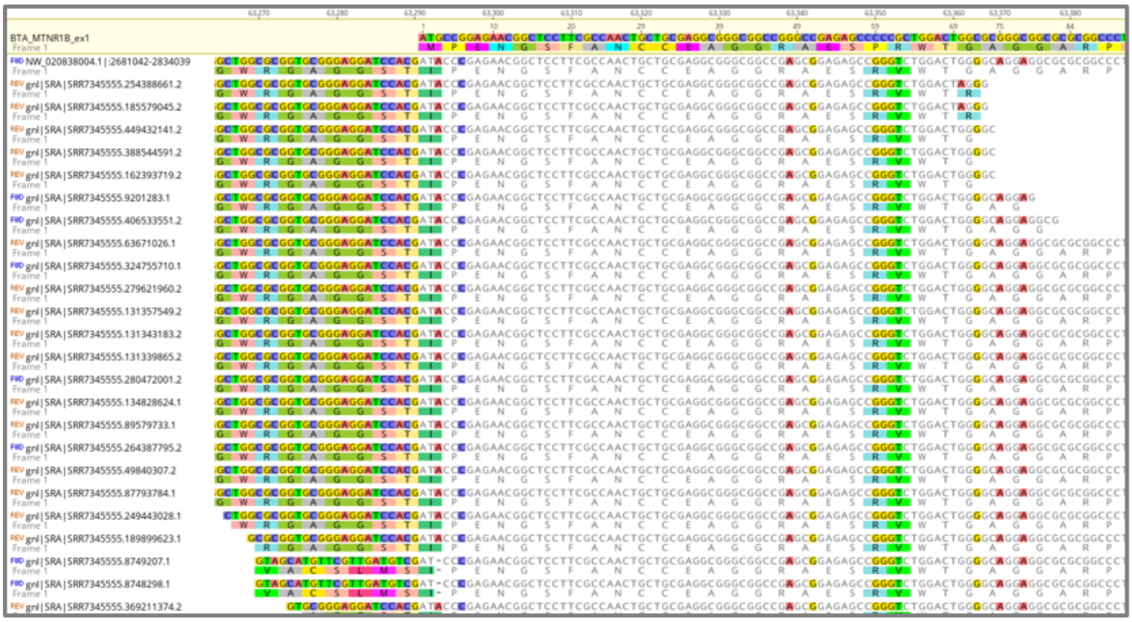


**Figure S17:** SRA validation for loss of start codon in exon 1 of the *Mntr1b* gene from *L. obliquidens*.

```
                                10        20        30        40        50        60        70        80        90       100
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
B. taurus NM_001206907   GTAACCTGTTCTTGGTGAGTCTGGCGTTCGCTGACCTGGCCGTAGCCTTGTACCCCTACCCGCTGAGCCTCGTGGCCATCTTCCACGACGGCTGGGCCCT
H. sapiens NM_005959     ....TT...................A..G..........TG..G.....C..............A.T................T.T..............
S. chinensis             ....TT...................A.CG.........------.......G............A.T....T............................
D. leucas                ..G.TT...................A.CG.........------.......G............A.T....T.................A..........
L. vexillifer            ....TT...................A..G.........------.......G............A.T....T............................
N. asiaeorientalis       ....TT...................A.CG.........------.......G............A.T....T.................A..........
P. catodon               ....TT...................A..G.........------....................A.T....T.............A...A..........
B. acutorostrata         ....TT...................A..G.........------....................A.T....T.............T..............
B. bonaerensis           ....TT...................A..G.........------....................A.T....T.............T..............
E. robustus              ....TT...................A..G.........------....................A.T....T............................
B. mysticetus            ....TT...................A..GAA.......------....................A.T..C.T............................

                               110       120       130       140       150       160       170       180       190       200
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
B. taurus NM_001206907   -GGGGGAGGTGCACTGCAAGGCCAGCGCCTTCGTGATGGGTCTGAGCGTGGTCGGCTC------CGTCTTCAACATCACCGCCATCGCCGTCGACCGCTA
H. sapiens NM_005959     -........A.....................T........C........CA.......------T........T.....T.........A.TA.......
S. chinensis             G...........G..................T........C........CA..A....------.......G....T.....TG.....A.TA.....--
D. leucas                -........C..G.............C....T.................CA.......------.......G.......T...G.T...A.TA..T....
L. vexillifer            -........C..G......AC..........T........C........CA.......------T......G...........G.....C.TA.......
N. asiaeorientalis       -........C.TG.............C....T..TG.............CA.......CGTCTT.......G.......T.....T...A.TA..T....
P. catodon               -........C..G...........A......T........C........CA.......------.......G.................A.TA.......
B. acutorostrata         -........C..G..................T........C........CA.T.....------.......G.................A.TA.......
B. bonaerensis           -........C..G..................T........C........CA.T.....------.......G.................A.TA.......
E. robustus              -........C..G..................T........C........CA.T.....------.......G.................A.TA.......
B. mysticetus            -...........G..................T........C........CA.......------.A.....G.............T...A.TA.A.A...

                               210       220       230       240       250       260       270       280       290       300
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|  (…)
B. taurus NM_001206907   CTGC-TACGTCTGCCGCAGCGTGACCTACCACCGCCTCTGCCGCCGCCGGCACGCCGCCGTCTACGTTGGCCTCGTCTGGCTGCTCACCCTGCTGGTCCT
H. sapiens NM_005959     ....-...A......A....A..G..........AA...A...G...T.....A..C.TC.GC..A.CT.....A.......C......G..G...C.T.
S. chinensis             ....-..........A.................T.A.....T...A.T.....A..C..C.....A.CT....T.....T..C.............-...
D. leucas                ....-..........A...T...............A.....T...A.T.....A..C..C.....A.CT....T.....T..C.............-...
L. vexillifer            ....-..........A...................A.....T...A.T.....A..C..C.....A.CT....T.....T..C.............-...
N. asiaeorientalis       ....-..........A.................T.A.....T...A.T.....A..C..C.....A.CT....T.....T..C.............-...
P. catodon               ....T..........A............A......A...A.T...A.T.....A..C..C.....A.CT....T........C......T......-...
B. acutorostrata         ....-..........A...................A...A.T...A.T..T..A..C..C.....A.CT....T........C.............-...
B. bonaerensis           ....-..........A...................A...A.T...A.T..T..A..C..C.....A.CT....T........C.............-...
E. robustus              ....-.G........A...................A...A.T...A.T..--------------------------------------------------
B. mysticetus            ....-..........A................A..A...A.T...A.T.....A..C..C.....A.CT....T...C....C.............-...
```

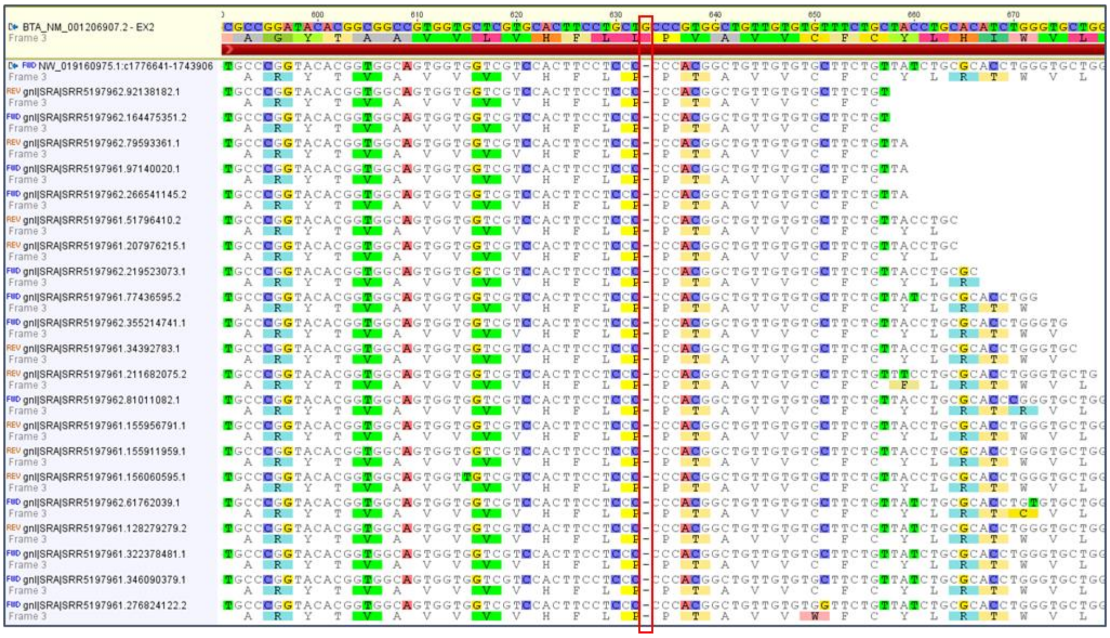
**Figure S18:** Partial multiple alignment of the predicted exon 2 of *Mtnr1b* in the listed species. Conserved nucleotide deletion, validated with SRA (when available), is represented in the corresponding position with a black arrow.

***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)

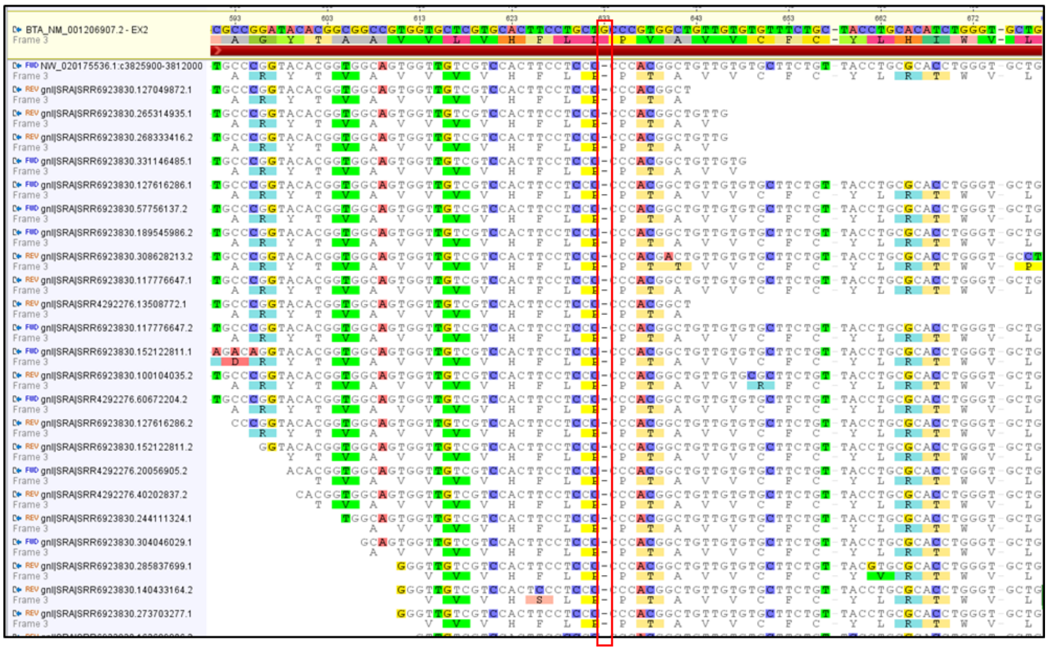


***Neophocaena asiaeorientalis*** conserved 1 nt deletion exon 2

SRA searched

**SRR6923630**- Northwestern Polytechnical University 2018-04-08 **Sample ID:** SAMN08512128

**SRR4292276**- Beijing Genome Institute 2017-09-23 **Sample ID:** SAMN05818630

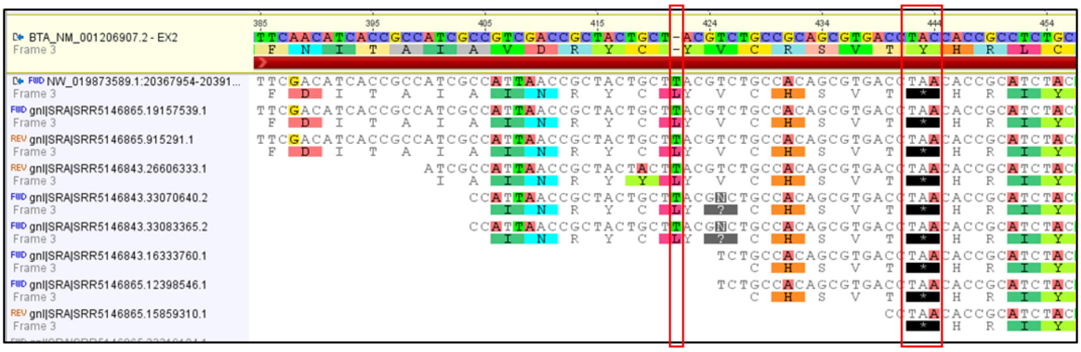


***Physeter catodon*** 1 nt deletion and stop codon exon 2

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



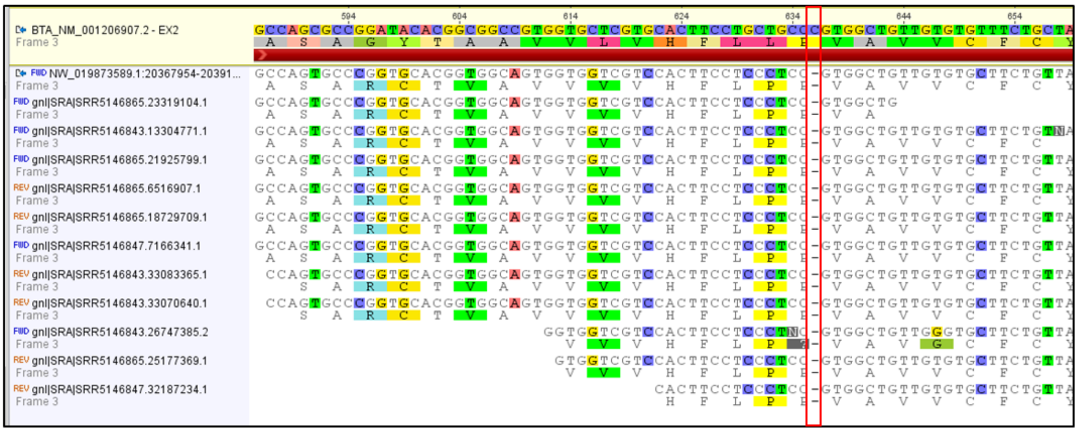
***Physeter catodon*** conserved 1 nt deletion exon 2

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

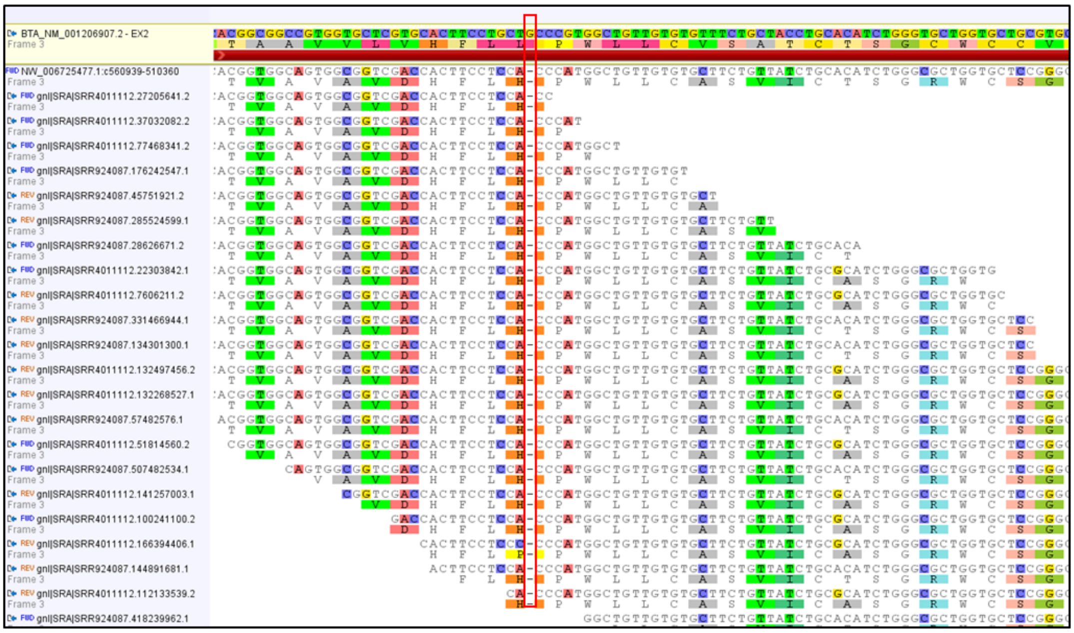


***Balaenoptera acutorostrata*** Conserved 1 nt deletion exon 2

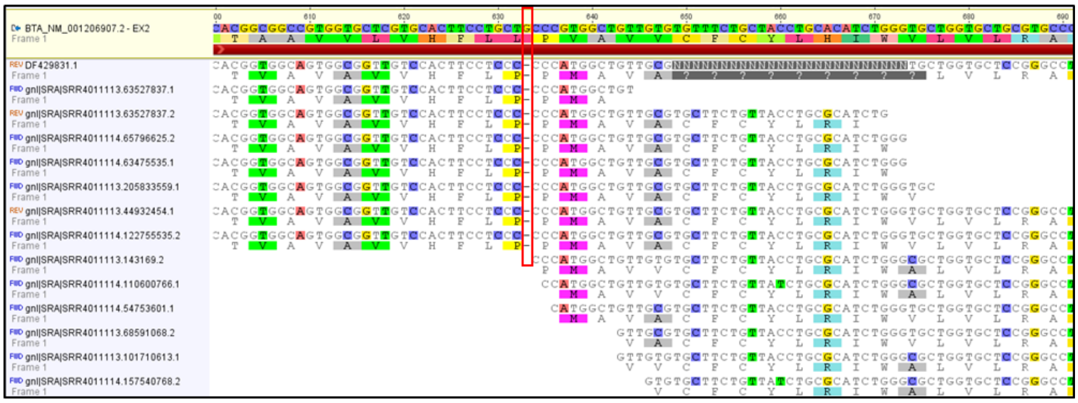
SRA searched

**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)



***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



***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)

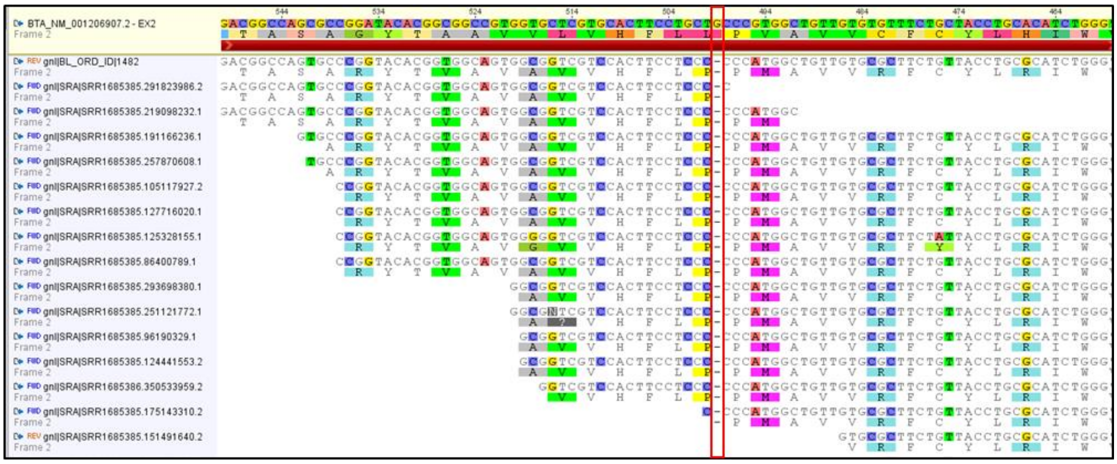


**Figure S19:** SRA validation for conserved nucleotide deletion in exon 2 of *Mtnr1b* from *D. leucas*, *N. asiaeorientalis*, *P. catodon*, *B. acutorostrata*, *B. bonaerensis* and *B. mysticetus*.

***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)

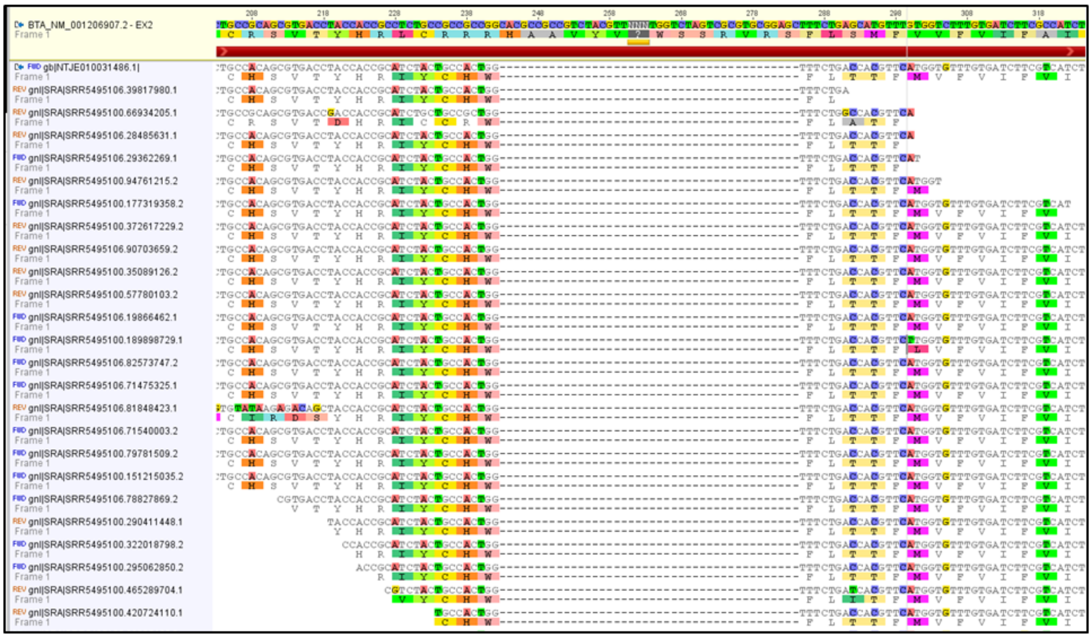


**Figure S20:** SRA validation for nucleotide deletion in exon 2 of *Mtnr1b* from *E. robustus*.

***Hippopotamus amphibius*** *Asmt* Exon 1 assembly using SRA

SRA searched

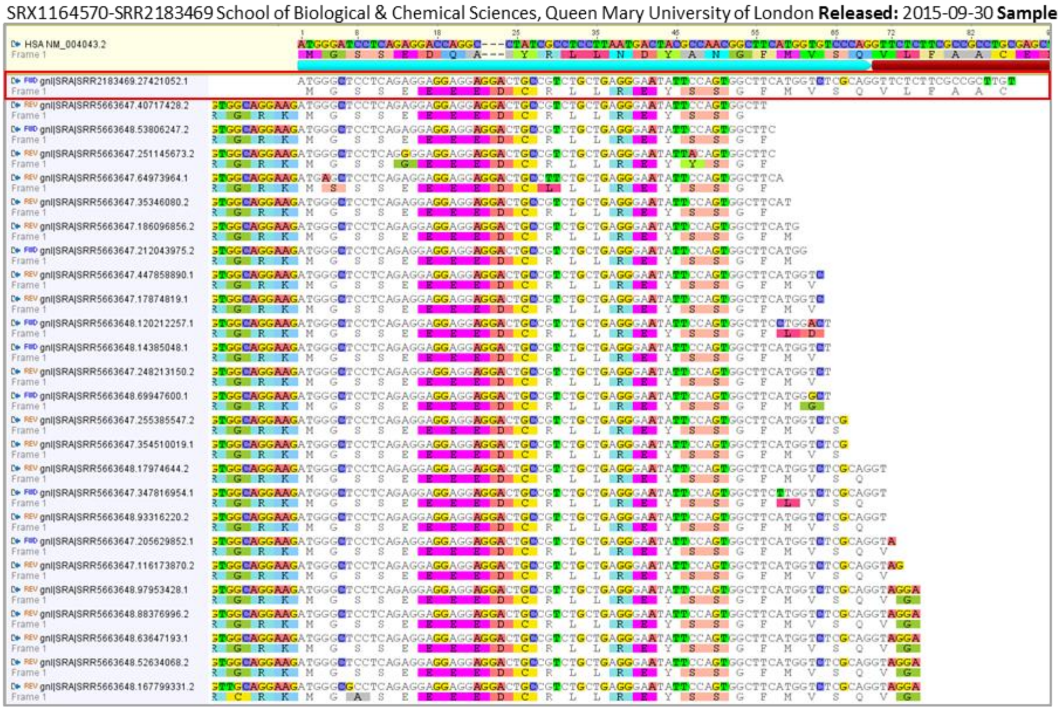
**WGS**

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



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**

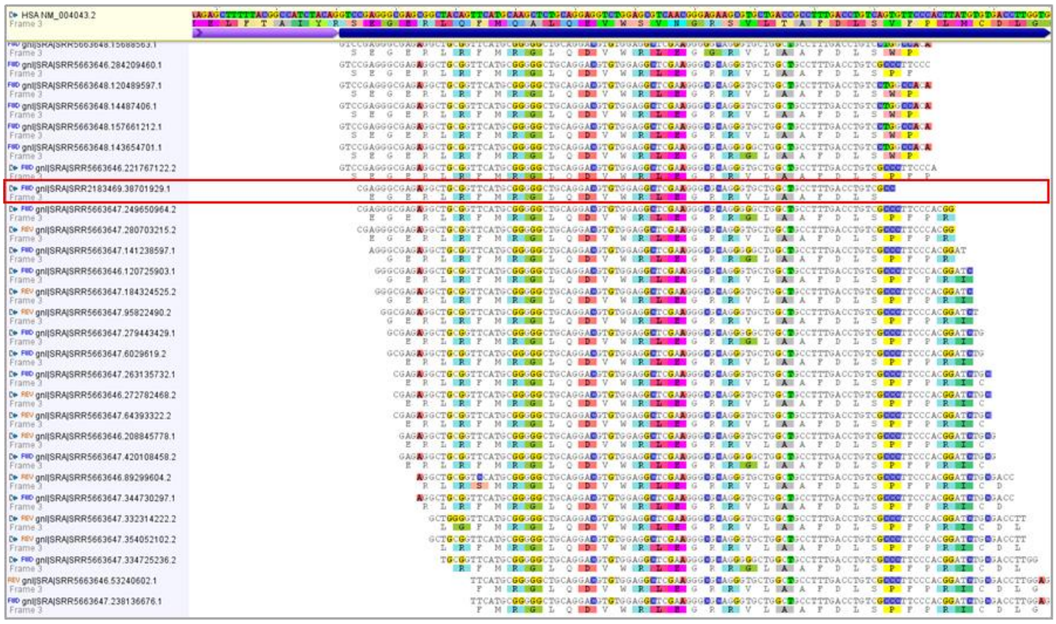
SRR5663648 -Senckenberg Gesellschaft fuer Naturforschung **Released:** 2018-04-05 **Sample:** SAMN07206988

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



Red box highlights RNA-Seq read mapping to exon 5

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

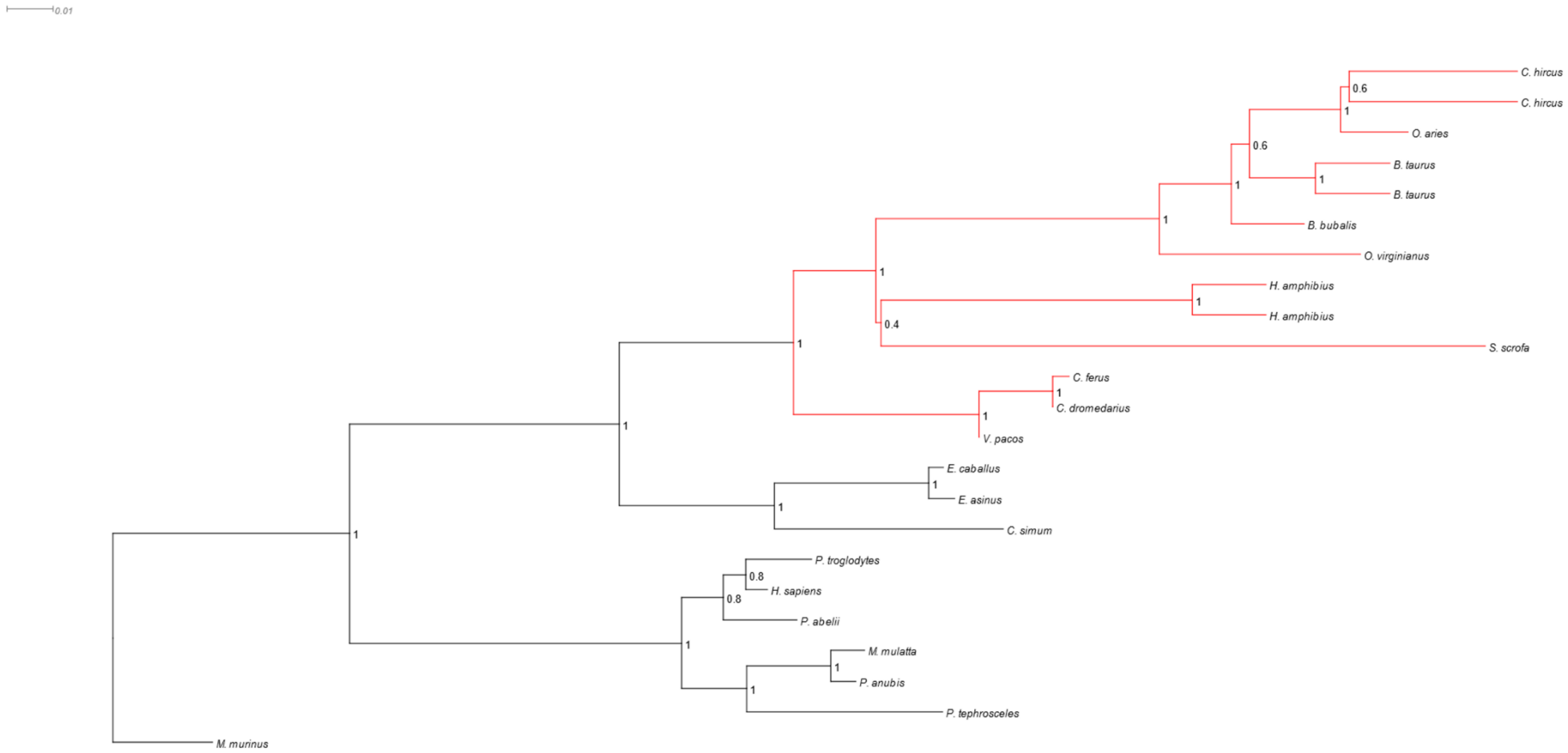


**Figure S22:** Maximum likelihood phylogenetic analysis of Perissodactyla, Artiodactyla and Primate *Aanat* genes. Node values represent branch support using the aBayes algorithm. In red the clade containing the *H. amphibius*.

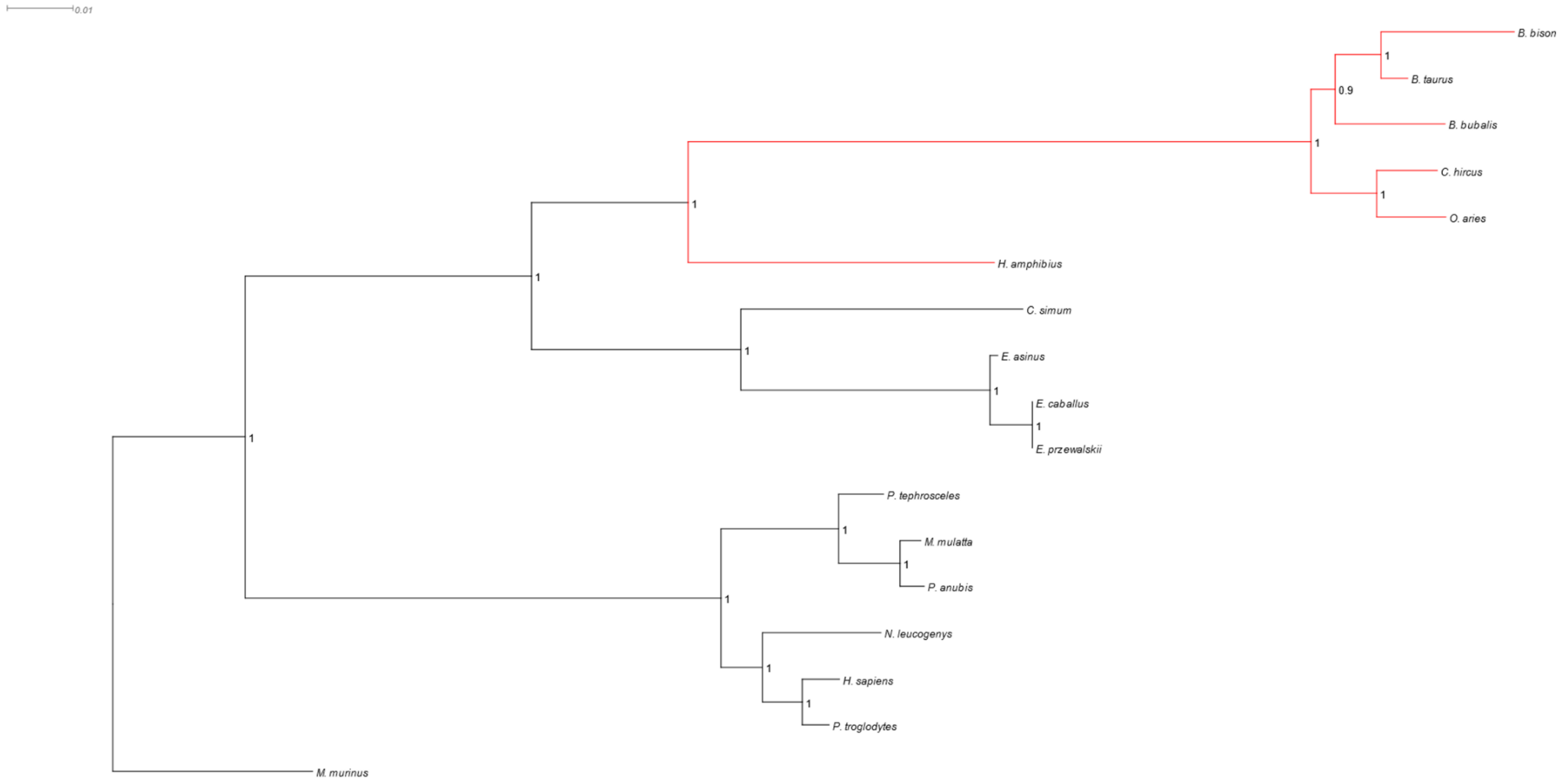


**Figure S23:** Maximum likelihood phylogenetic analysis of Perissodactyla, Artiodactyla and Primate *Mtnr1a* genes. Node values represent branch support using the aBayes algorithm. In red the clade containing the *H. amphibius*.

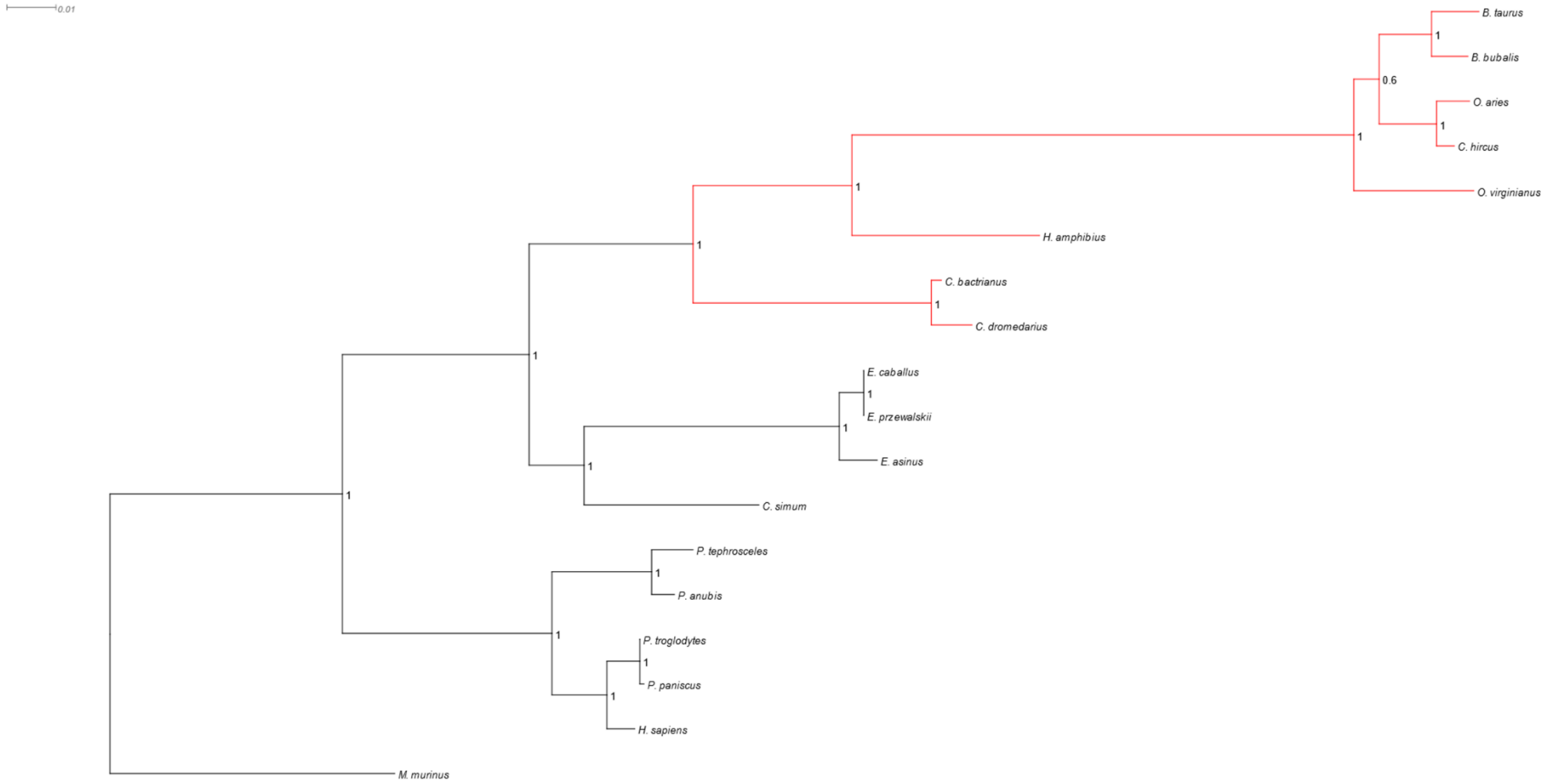


**Figure S24:** Maximum likelihood phylogenetic analysis of Perissodactyla, Artiodactyla and Primate *Mtnr1b* genes. Node values represent branch support using the aBayes algorithm. In red the clade containing the *H. amphibius*.

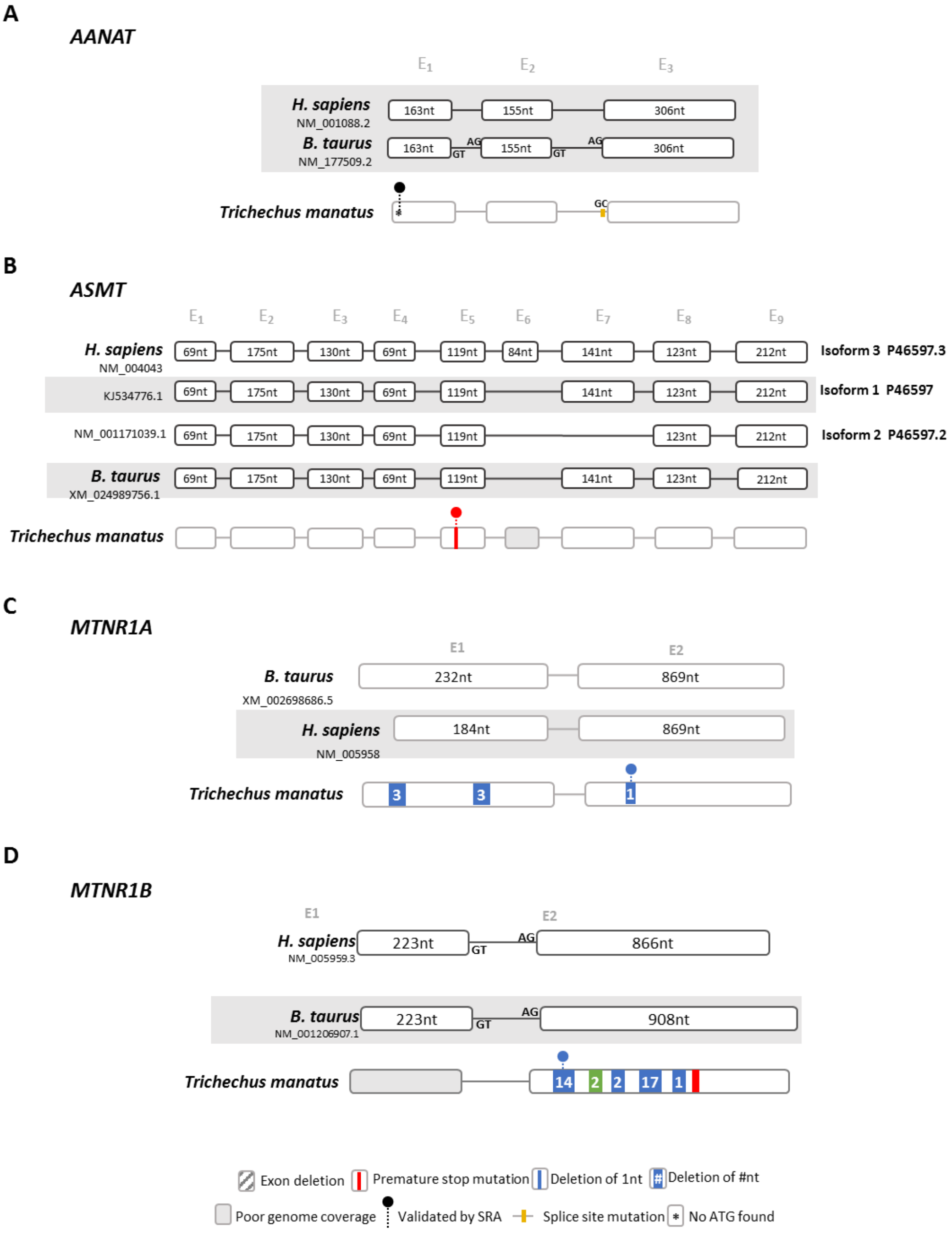


**Figure S25:** *T. manatus latirostris* gene annotations. Schematic representation of the gene structure of human and B. taurus *Aanat* (A), *Asmt* (B), *Mtnr1a* (C) and *Mtnr1b* (D) genes, each box represents an exon and lines represent intronic region, and sSchematic representation of the corresponding 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

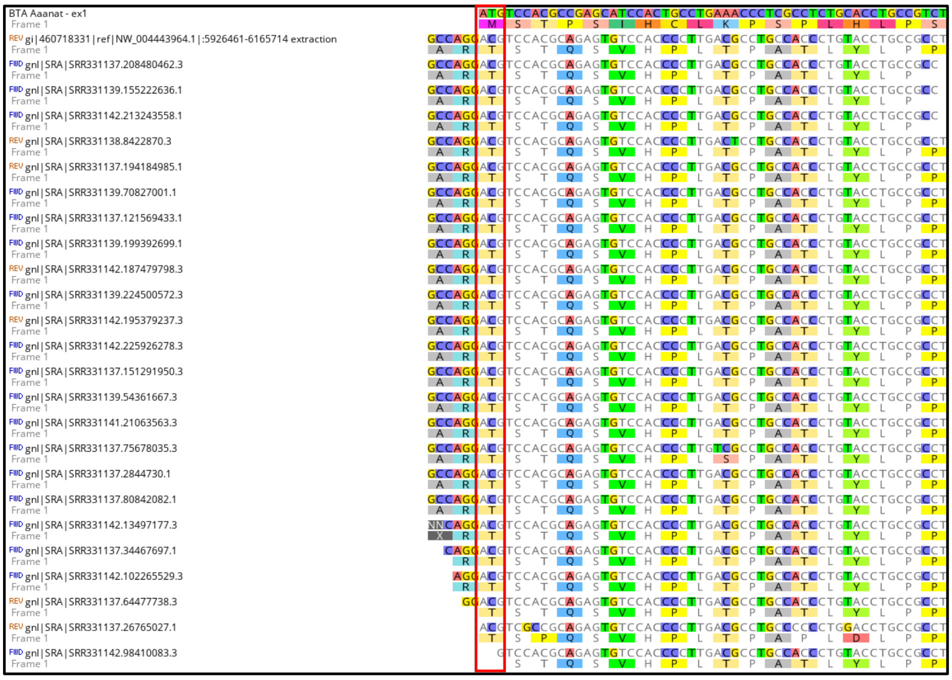


**Figure S26:** SRA validation for loss of start codon in exon 1 of the *Aanat* gene from *T. manatus latirostris.*

***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

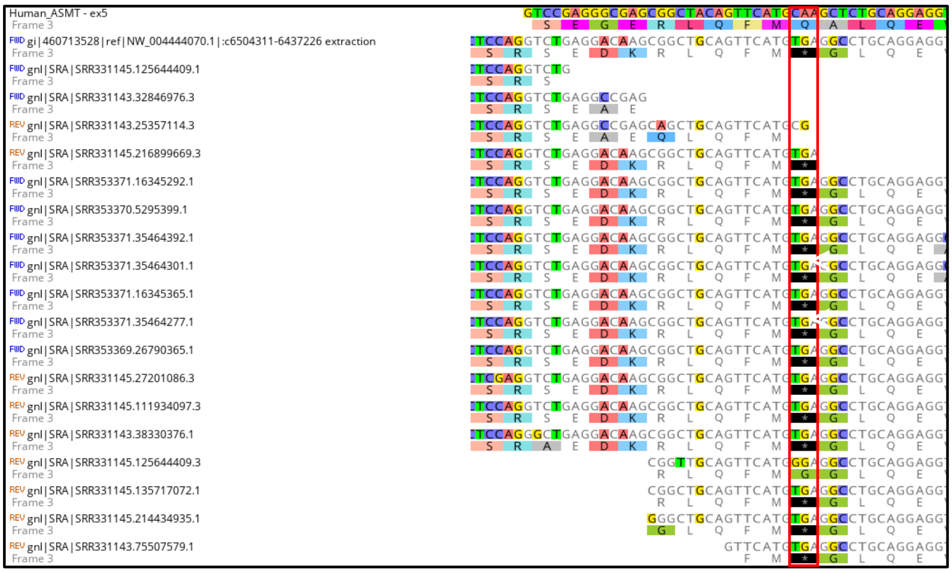


**Figure S27:** SRA validation for premature stop codon in exon 5 of the *Asmt* gene from *T. manatus latirostris.*

***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

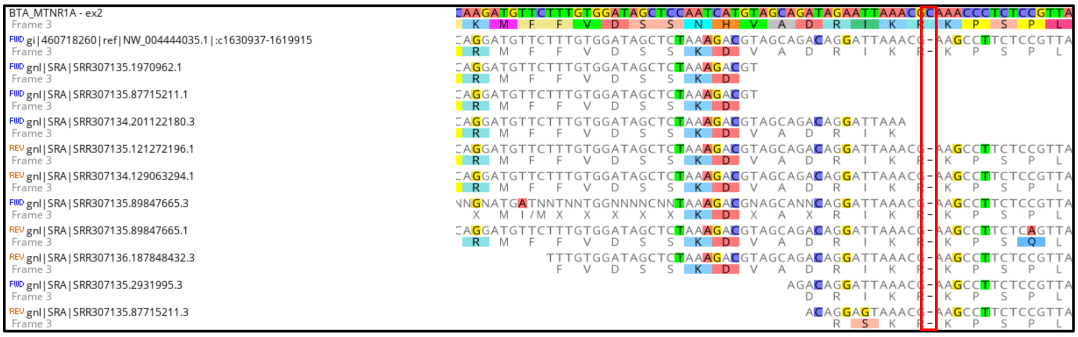


**Figure S28:** SRA validation for single nucleotide insertion in exon 2 of the *Mtnr1a* gene from *T. manatus latirostris.*

***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

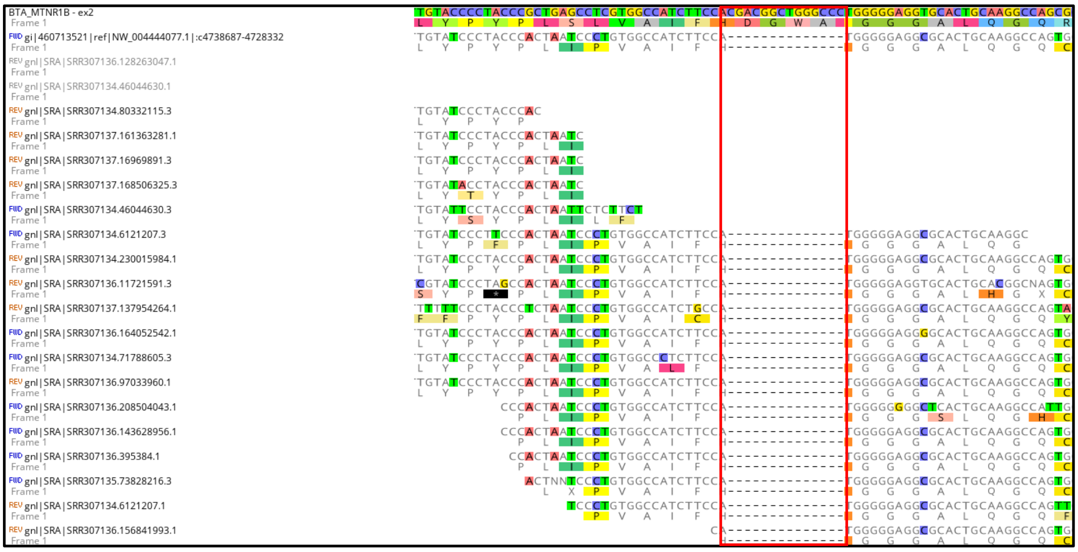


**Figure S29:** SRA validation for frameshift mutation in exon 2 of the *Mtnr1b* gene from *T.manatus latirostris.*

**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*.

| **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 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 transcript 43218 44549 0.28 - . g2.t1

NKPW01005924.1 AUGUSTUS stop_codon 43218 43220 . - 0
transcript_id "g2.t1"; gene_id "g2";

NKPW01005924.1 AUGUSTUS terminal 43218 43503 0.38 - 1
transcript_id "g2.t1"; gene_id "g2";

NKPW01005924.1 AUGUSTUS internal 43916 44066 0.37 - 2
transcript_id "g2.t1"; gene_id "g2";

NKPW01005924.1 AUGUSTUS initial 44387 44549 0.8 - 0 transcript_id
"g2.t1"; gene_id "g2";

NKPW01005924.1 AUGUSTUS CDS 43218 43503 0.38 - 1 transcript_id
"g2.t1"; gene_id "g2";

NKPW01005924.1 AUGUSTUS CDS 43916 44066 0.37 - 2 transcript_id
"g2.t1"; gene_id "g2";

NKPW01005924.1 AUGUSTUS CDS 44387 44549 0.8 - 0 transcript_id
"g2.t1"; gene_id "g2";

NKPW01005924.1 AUGUSTUS start_codon 44547 44549 . - 0
transcript_id "g2.t1"; gene_id "g2";

coding sequence =
[atgtccacacagaccatccaatacatgaagcctgtggctctgcacctgccacctgggatcccagagtcgccaagccgccagcggcgccataca
ctccccgccaatgagttccgctgcctcaccccagaggacgctgctgacgtgtttgagattgagcgagaggccttcatctctgtctcgggcgcctgc
cccctgcaactggaccaggtccagcacttcctgaccgtgtgtcccgagctgtccctgggctgcttcgtggagggccgcctcgtggccttcatcatt
ggctccctgtgggacaaggagagactcaggcccgggggccacacagcccacctgcacttgctggccgtgcgccgccccttccggcgtctggg
caacggttccaccctgctctggcgctctctgcaccacctggatggcaagccggccgtgcgccggacctcgctcatgtgcgaggacccattggtgc
ccttctaccagaggttcggcttccaccctgtgggcccgtgtgccgtgaccgtgggctccctcaccttcacggagatggagcgctccccgtggggc
catgcctccctgcgcaggaacagtgacagctga]

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	AUGUSTUS	stop_codon	54868	54870	.	-	0
	transcript_id "g3.t1"; gene_id "g3";

NKPW01005924.1	AUGUSTUS	terminal	54868	55173	0.93	-	0
	transcript_id "g3.t1"; gene_id "g3";

NKPW01005924.1	AUGUSTUS	internal	55567	55721	0.97	-	2
	transcript_id "g3.t1"; gene_id "g3";

NKPW01005924.1	AUGUSTUS	initial	56042	56204	0.95	-	0	transcript_id
"g3.t1"; gene_id "g3";

NKPW01005924.1	AUGUSTUS	CDS	54868	55173	0.93	-	0	transcript_id
"g3.t1"; gene_id "g3";

NKPW01005924.1	AUGUSTUS	CDS	55567	55721	0.97	-	2	transcript_id
"g3.t1"; gene_id "g3";

NKPW01005924.1	AUGUSTUS	CDS	56042	56204	0.95	-	0	transcript_id
"g3.t1"; gene_id "g3";

NKPW01005924.1	AUGUSTUS	start_codon	56202	56204	.	-	0
	transcript_id "g3.t1"; gene_id "g3";

coding sequence =
[atgtccacacagaccatccaatacatgaagcctgtggctctgcacctgccacctgggatcccagagtcgccaagccgccagcggcgccataca
ctccccgccaatgagttccgctgcctcaccccagaggacgctgctgacgtgtttgagattgagcgagaggccttcatctctgtctcgggcgtctgcc
ccctgcaactggaccaggtccagcacttcctgaccgtgtgtcccgagctgtccctgggctgcttcgtggagggccgcctcgtggccttcatcattg
gctccctgtgggacaaggagagactcactcaggagtcactgacactacacaggcccgggggccacacagcccacctgcacttgctggccgtgc
accgcaccttccggcagctggggaagggctccaccctgctctggcgttacctgcaccatctggatggccagccggccgtgcgccgggccgtgc
tcatgtgcgaggacccattggtgcccttctaccagaggttcggcttccaccctgttggcctgtgtgccgcgaccgtgggctccctcaccttcacgga
gatgcaatgctccctgaggggccacgcctccctgcgcagaaacagtgacagctga]

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 g1

NKPW01072562.1_NKPW01009466.1       AUGUSTUS  gene    770       16959  0.32     +        .
        g1

NKPW01072562.1_NKPW01009466.1       AUGUSTUS  transcript        770       16959  0.32     +
        .         g1.t1

NKPW01072562.1_NKPW01009466.1       AUGUSTUS  start_codon       770       772    .        +
        0         transcript_id "g1.t1"; gene_id "g1";

NKPW01072562.1_NKPW01009466.1       AUGUSTUS  initial 770       998    1        +        0
        transcript_id "g1.t1"; gene_id "g1";

NKPW01072562.1_NKPW01009466.1       AUGUSTUS  terminal          16091  16959  0.32     +
        2         transcript_id "g1.t1"; gene_id "g1";

NKPW01072562.1_NKPW01009466.1       AUGUSTUS  CDS     770       998    1        +        0
        transcript_id "g1.t1"; gene_id "g1";

NKPW01072562.1_NKPW01009466.1       AUGUSTUS  CDS     16091     16959  0.32     +        2
        transcript_id "g1.t1"; gene_id "g1";

NKPW01072562.1_NKPW01009466.1       AUGUSTUS  stop_codon        16957  16959  .        +
        0         transcript_id "g1.t1"; gene_id "g1";

coding sequence =
[atggctggggggccgtggggcgcgccgggaggggcctccaagggcaacggcagcgcgctgctcaatgcctcgcagcaggcgcctggcag
cggggagggcgcgccgtcgcgcccctcttggctggtcaccacgctcgccctcatcctcatcttcaccatcgtggtggacatcctgggcaacctcct
ggtcatcctgtccgtgtatcggaacaagaagctgaggaacgcaggaaatatatttgtggtgagcctagcagttgcagacctgctggtggccgtgta
tccataccccttggccctgacatctatcttcaacaatgggtggagcctgggctacctgcactgccagattagtggcttcctgatgggcttgagtgtcat
tggctccgtattcaatatcacggggattgccatcaaccgctattgctacatctgccacagtctccagtacgacaggttgtacagcaacaagaattccc
tctgctatgtgttcctgatatggatgttgacacttgtggcaatcgtgcccaacttgtgtattggaaccctgcggtacgacccgaggatctattcctgtac
attcacacagtccatcagttccgcgtacacgatagccgtggtggttttccatttcatggttcctatggtcatagtaatcttctgttacctaagaatctgggt
cctggttcttcaggtcagatggagggtgaaacctgacaacaaacccaaactgaaaccacaggacttcaggaattttgtcaccatgtttgtggtttttgt
actttttgccatttgctgggcgcctctaaacttcattggtcttgctgtggcctcagaccccgccagcatggtatccaggatcccagagtggctgtttgtg
gccagttactatatggcgtatttcaacagctgcctcaacgcaattatatatggactactgaaccaaaatttcaggcaggaatacaggagaattatagtc
tcactgtgcacagccaagatgttctttgtggatagctctaacaatgtcgcagatagaatcaaatgcaaaccctctccattaataaccaaccataaccta
ataaaggtggactccgtttaa]

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

NKPW01013551.1    AUGUSTUS   gene    25975  40822  1       -       .       g1

NKPW01013551.1    AUGUSTUS   transcript      25975  40822  1       -       .       g1.t1

NKPW01013551.1    AUGUSTUS   stop_codon      25975  25977  .       -       0
        transcript_id "g1.t1"; gene_id "g1";

NKPW01013551.1    AUGUSTUS   terminal        25975  26813  1       -       2
        transcript_id "g1.t1"; gene_id "g1";

NKPW01013551.1    AUGUSTUS   initial  40600  40822  1       -       0       transcript_id
"g1.t1"; gene_id "g1";

NKPW01013551.1    AUGUSTUS   CDS     25975  26813  1       -       2       transcript_id
"g1.t1"; gene_id "g1";

NKPW01013551.1    AUGUSTUS   CDS     40600  40822  1       -       0       transcript_id
"g1.t1"; gene_id "g1";

NKPW01013551.1    AUGUSTUS   start_codon     40820  40822  .       -       0
        transcript_id "g1.t1"; gene_id "g1";

coding sequence =
[atgcccgagaaccgctccttcgccaactgctgcgaggcgggcggccgagccgagagcccaggctggactggggcgggcggcgcgcggcc
ctccgggaccccccggcctccctgggtggcgcccgcgctgtccgccgtgctcatcgtcaccaccgcggtggacatcgtgggcaacctcctggtc
atcctctcggtgctcgggaaccgcaagctccggaacgcaggtaatttgttcttggtgagtctggcattggctgacctggcggtggccttgtacccct
acccgctaatcctcgtcgccatcttccacaacggctgggccctgggggaggcgcactgcaaggccagcgccttcgtgatgggtctgagcgtcat
cggctctgtcttcaacatcaccgccatcgccattaaccgctactgctacgtctgccgcagcgtgacctaccaccgcatctaccgccactggcacac
cgccctgtacatctgccttgtctggctgctcaccctgctggccgtggtacccaacttcttcttggggtccctggagtatgaccctcgcatctactcttgc
accttcatccagacggccagcgccgggtacacggtggccgtggtggtggtccacttcctcctccccatggctatcgtgtgctcctgttacctgcgca
tctgggtgctggtgctccgggcccgcaggaaggtcaaggcggagagcaagccgtgcctgtggtccaggaacgtccggagcttcctgaccatgtt
cgtggtgtttgtgatcttcgccatctgctgggcgcccctgaactgcattggcctcgctgtggccatcgacccagaagaagtggctccccgggtccc
agaggggctctttgtcacgagctacttcctggcttatttcaacagctgtcttaatgccatcatctacgggcttctgaaccagaacttccgcagggaata
caagaagatcgtctctgccctctggaacccacggcgctgccagcaggactcttccaagggcagccaggcggagggccagagacccaagctc
cccgccgtggttaa]

protein sequence =
[MPENRSFANCCEAGGRAESPGWTGAGGARPSGTPRPPWVAPALSAVLIVTTAVDIVGNLLV
ILSVLGNRKLRNAGNLFLVSLALADLAVALYPYPLILVAIFHNGWALGEAHCKASAFVMGL
SVIGSVFNITAIAINRYCYVCRSVTYHRIYRHWHTALYICLVWLLTLLAVVPNFFLGSLEYDP
RIYSCTFIQTASAGYTVAVVVVHFLLPMAIVCSCYLRIWVLVLRARRKVKAESKPCLWSRNV
```

RSFLTMFVVFVIFAICWAPLNCIGLAVAIDPEEVAPRVPEGLFVTSYFLAYFNSCLNAIIYGLL
NQNFRREYKKIVSALWNPRRCQQDSSKGSQAEGPETQAPRRG]