

# MHCIIDQB Phylogenetic Tree of the Irrawaddy dolphin in Relation to Cetaceans

Sett Aung

2024-05-24

```
library(tidyverse)
library(openintro)
library(ggtree)
library(ape)
```

## Introduction

Species become endangered because of habitat loss and genetic variation. One of the species at the risk of extinction are the Irrawaddy dolphin (*Orcaella brevirostris*) listed as endangered by the International Union for Conservation of Nature (IUCN). (Minton, et al. 2017). The Irrawaddy dolphin are small cetacean species that inhabit throughout South-East Asia rivers, estuarine, and coastal environment distributing from North-East India, Myanmar, Thailand, Cambodia, Laos, Vietnam, Indonesia to the east of the Philippines. In coastal habitat, are often associated with freshwater discharges, and estuaries (Jackson Ricketts, et al. 2018). The freshwater subpopulation occurs in three large rivers which includes the Ayeyarwaddy (Irrawaddy) River in Myanmar, Mahakam River in Indonesia, and Mekong River in Cambodia and Southern Laos. This species is also found as a lacustrine subpopulation in lakes such as Chilika Lake in India, Songkhla Lagoon in Thailand, and Malampaya Sound in the Philippines (Beasley et al. 2002, Dai et al. 2021). The riverine subpopulation are considered critically endangered by IUCN and their long-term survival has become a conservative concern due to their fragmented distribution and threats from human activities such as electrofishing and entanglement in gillnets (Gilbert et al. 2009, Brownell et al. 2019), habitat loss from dam constructions in riverine populations (Brownell et al. 2017), contamination from pollutants (such as oil, pesticides, and industrial wastes) in the estuarine population (Schnitzler et al. 2021, Murphy et al. 2007), and vessel strikes in some areas (Thomas et al. 2019). With restricted dispersal, and possible population bottlenecks, some isolated subpopulations will be subjected to inbreeding resulting in limited gene flow (Frankham, 1995) which could lead to loss of genetic variation (Hamner, et al. 2012).

The major histocompatibility complex (MHC) genes have been widely studied to examine genetic variation and determine population health (Sommer, 2005). MHC class II genes are highly polymorphic, especially the peptide binding region (PBR) that is involved with recognition and binding with pathogen antigens that initiate T-lymphocyte immune response (Santillan et al. 2016; Caballero et al. 2010). Past studies have hypothesized that MHC polymorphism could be maintained by over-dominant selection or a balancing selection that favors heterozygosity or a pathogen mediated selection (Sommer, 2005; Hughes, 1995). It has not been clear which hypothesis supports the polymorphism of MHC.

Previous research by Hayashi et al. 2003, has found only one version of DQB gene (and only one DQB allele) because of the sample size (n=1). My goal to study genetic variation at the DQB locus in Irrawaddy dolphins across the South-East Asia distribution range is to determine the Irrawaddy dolphin population health for conservation and provide more information about the mechanism in maintaining variability of MHC genes. Therefore, I propose to test the following goals:

## Experimental Design

With thanks to Dr. Michael McGowen, the marine mammal curator from Smithsonian National Museum of Natural History, we obtained Irrawaddy dolphin teeth and bone samples from recently stranded individuals from the Department of Marine and Coastal Resources (DMCR) in Klaeng, Rayong, Thailand (n=11) and the MareCet Research Organization (MareCet) in Selangor, Malaysia (n=2). We also obtained tissue sample from various sources including the Southwest Fisheries Science Center's (SWFSC) Marine Mammal and Turtle Molecular Research Tissue Sample Collection (MMASTR) in La Jolla, CA, USA (n=29), the DMCR (n=11), and the MareCet (n=6). Since the species is under risk of extinction, we obtained historic samples originally collected between 1875-1972 which are osteological samples (n=16) and bone samples from *Orcaella heinsolani* (n=2) stored in eight museum collections in the United States and Europe. Unfortunately, we were not able to obtain modern samples in regions such as Sumatra (Indonesia), Ayeyarwaddy River (Myanmar), Chilika Lake (India), and the northern coast of Borneo.

All 77 samples were DNA extracted, Illumina NGS sequenced, and will be analyzed to reveal potential additional DQB alleles. From this experiment, we can identify additional DQB alleles that may be useful to recover the populations.

## Aim 2 Construct a phylogenetic tree to compare the DQB alleles from this study with the DQB alleles from cetaceans.

Rationale: Mechanisms that maintain MHC genetic diversity remain an open question in biology. There have been theories hypothesizing the maintenance of high variability of the MHC genes. This aim will be done to enrich the understanding of the evolution and relationship of the DQB alleles among other cetaceans.

Null hypothesis: There is no relationship between the MHCIIDQB Alleles of Irrawaddy dolphins with some of the cetaceans. Alternative hypothesis: There is a relationship between the MHCIIDQB Alleles of Irrawaddy dolphins with some of the cetaceans.

Methods: To analyze the relationship of the DQB alleles with few other cetaceans, we will compile Cetacean MHCII DQB sequences from GenBank and align them to the DQB alleles from Irrawaddy dolphin and input the phylogenetic tree from IQ Tool-CIPRES using R to make the tree better. The data from this approach will help us visualize which group of cetaceans has high similarity with the DQB alleles of Irrawaddy dolphins. This information can clarify more about the maintenance of high variability of MHC genes over time.

The data and script is stored in github repository <https://github.com/saung3/Orcaella-brevirostris/tree/main/Scripts>.

## Using ggtree

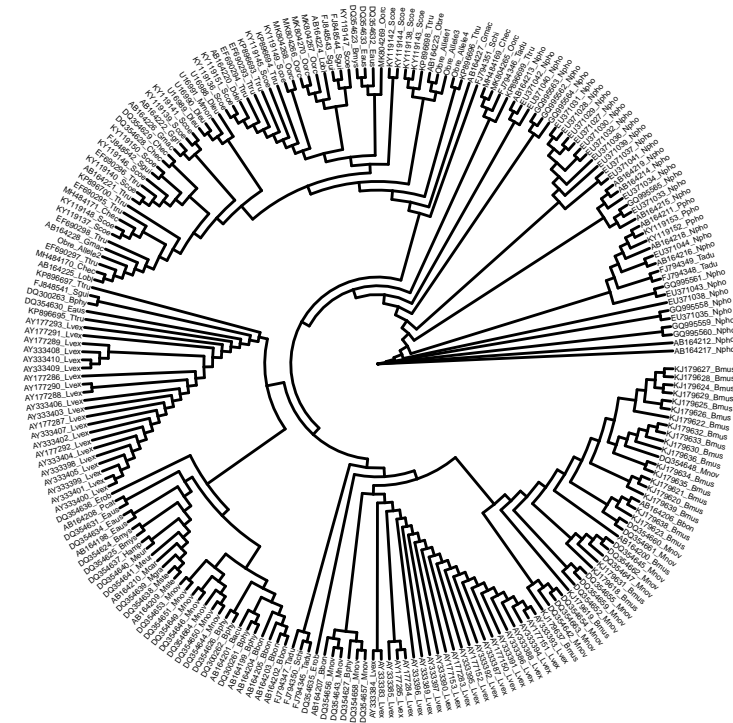
```
tree <- read.tree("../Data/output.bionj")

# Ladderizing the tree (Neighborjoining tree)
tree <- ladderize(tree, right = F)

# Importing NJ tree from CIPRES using IQ Tool
tree_1 <- ggtree(tree, branch.length='none', layout='circular') +
  geom_tiplab(size = 1) +
  ggtitle(
    "Neighbor-joining tree - Comparing DQB Alleles of Irrawaddy
      dolphins to DQB Alleles of cetaceans ")

tree_1
```

## Neighbor-joining tree – Comparing DQB Alleles of Irrawaddy dolphins to DQB Alleles of cetaceans



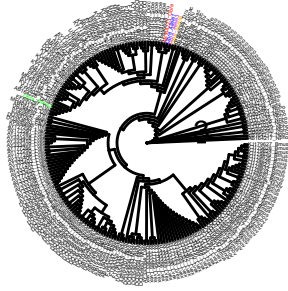
```
library(ggtree)
library(ape)
library(ggplot2)

# Assuming tree2 is your phylogenetic tree

# Convert the tree to a ggtree object with circular layout and no branch lengths
tree_gg_1 <- ggtree(tree, branch.length='none', layout='circular')

# Plot the tree with colored tip labels for specified groups
tree_gg_1 +
  geom_tiplab(size = 0.7, aes(color = ifelse(label %in% c("Obre_Allele1", "Obre_Allele2", "Obre_Allele3",
  scale_color_manual(values = c("red", "blue", "green", "purple", "orange"), na.value = "black") +
  ggtitle("NJ tree") +
  geom_treescale()
```

## NJ tree



```
ifelse(label %in% c("Obre_Allele1", "Obre_Allele2", "Obre_Allele3", ...
```

- AB164223\_Obre
- Obre\_Allele1
- Obre\_Allele2
- Obre\_Allele3
- Obre\_Allele4
- NA

## Using ggtree

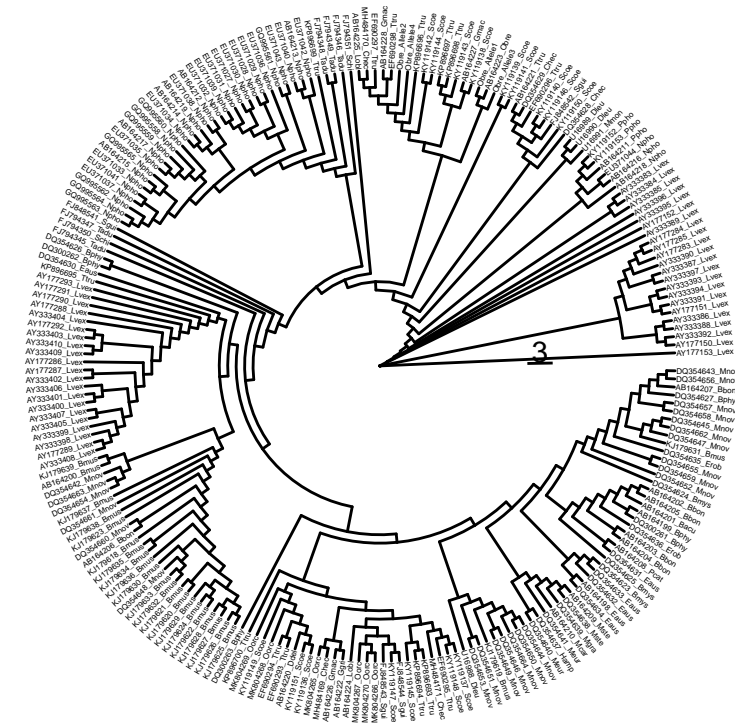
```
# reading contree output from CIPRES
tree2 <- read.tree("../Data/output.contree")
```

```
# Ladderizing
tree2 <- ladderize(tree2, right = F)
```

```
# Importing contree from CIPRES using IQ Tool
group = c("Obre_Allele1", "Obre_Allele2", "Obre_Allele3", "Obre_Allele4", "AB164223_Obre")
tree_2 <- ggtree(tree2, branch.length='none', layout='circular') +
  geom_tiplab(size = 1) +
  ggtitle("Bootstrap tree - Comparing DQB Alleles of Irrawaddy dolphins to
          DQB Alleles of cetaceans") +
  geom_treescale()

tree_2
```

## Bootstrap tree – Comparing DQB Alleles of Irrawaddy dolphins to DQB Alleles of cetaceans



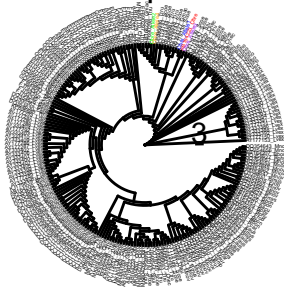
```
library(ggtree)
library(ape)
library(ggplot2)

# Assuming tree2 is your phylogenetic tree

# Convert the tree to a ggtree object with circular layout and no branch lengths
tree_gg_2 <- ggtree(tree2, branch.length='none', layout='circular')

# Plot the tree with colored tip labels for specified groups
tree_gg_2 +
  geom_tiplab(size = 0.7, aes(color = ifelse(label %in% c("Obre_Allele1", "Obre_Allele2", "Obre_Allele3",
  scale_color_manual(values = c("red", "blue", "green", "purple", "orange"), na.value = "black") +
  ggtitle("Bootstrap tree") +
  geom_treescale()
```

## Bootstrap tree



```
ifelse(label %in% c("Obre_Allele1", "Obre_Allele2", "Obre_Allele3", ...
```

- AB164223\_Obre
- Obre\_Allele1
- Obre\_Allele2
- Obre\_Allele3
- Obre\_Allele4
- NA

## Using ggtree

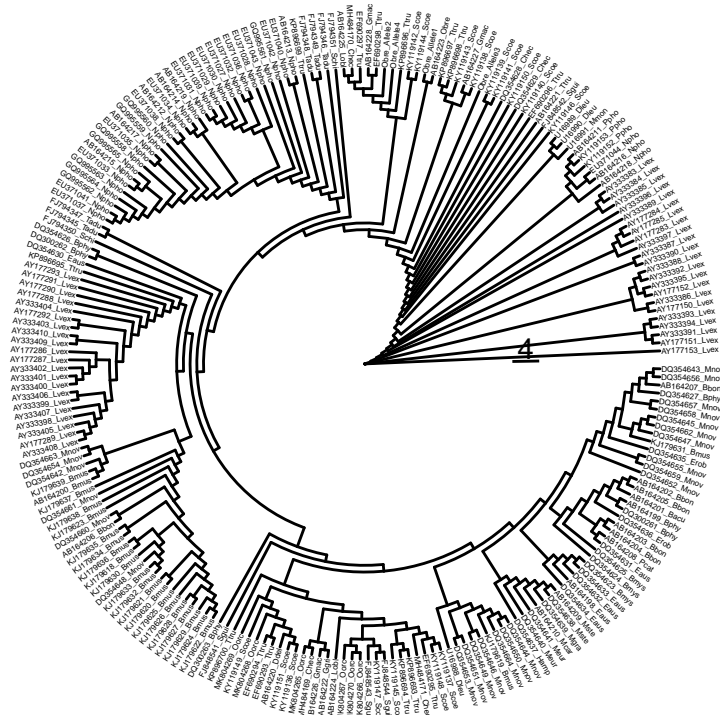
```
# reading ML treefile output from CIPRES
tree3 <- read.tree("../Data/output.treefile")
```

```
# Ladderizing
tree3 <- ladderize(tree3, right = F)
```

```
# Importing ML tree from CIPRES using IQ Tool
tree_3 <- ggtree(tree3, branch.length='none', layout='circular') +
  geom_tiplab(size = 1) +
  ggtitle("Maximum Likelihood tree - Comparing DQB Alleles of Irrawaddy
          dolphins to DQB Alleles of cetaceans") +
  geom_treescale()
```

```
tree_3
```

## Maximum Likelihood tree – Comparing DQB Alleles of Irrawaddy dolphins to DQB Alleles of cetaceans



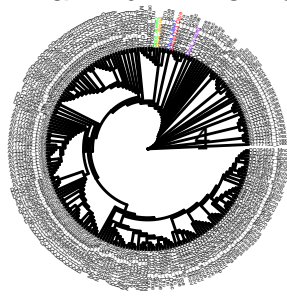
```
library(ggtree)
library(ape)
library(ggplot2)

# Assuming tree2 is your phylogenetic tree

# Convert the tree to a ggtree object with circular layout and no branch lengths
tree_gg_3 <- ggtree(tree3, branch.length='none', layout='circular')

# Plot the tree with colored tip labels for specified groups
tree_gg_3 +
  geom_tiplab(size = 0.7, aes(color = ifelse(label %in% c("Obre_Allele1", "Obre_Allele2", "Obre_Allele3",
  scale_color_manual(values = c("red", "blue", "green", "purple", "orange"), na.value = "black") +
  ggtitle("Maximum Likelihood tree") +
  geom_treescale()
```

## Maximum Likelihood tree



```
ifelse(label %in% c("Obre_Allele1", "Obre_Allele2", "Obre_Allele3", ...
```

- AB164223\_Obre
- Obre\_Allele1
- Obre\_Allele2
- Obre\_Allele3
- Obre\_Allele4
- NA

Overall the tree outputs show different result. The NJ Tree shows that Allele 1,3,4 shares the same clade along with the allele found from Hayashi et al. 2008. Bootstrap tree and maximum likelihood shows that Allele 1, 3 shares the same clade with the allele found from previous research and Allele 2,4 shares the same clade. Both the bootstrap and maximum likelihood tree shows that there is trans-species polymorphism for MHCIIIDQB alleles since the alleles of the same species are not cluster together but in separate clade. Therefore the results indicates the alternative hypothesis.

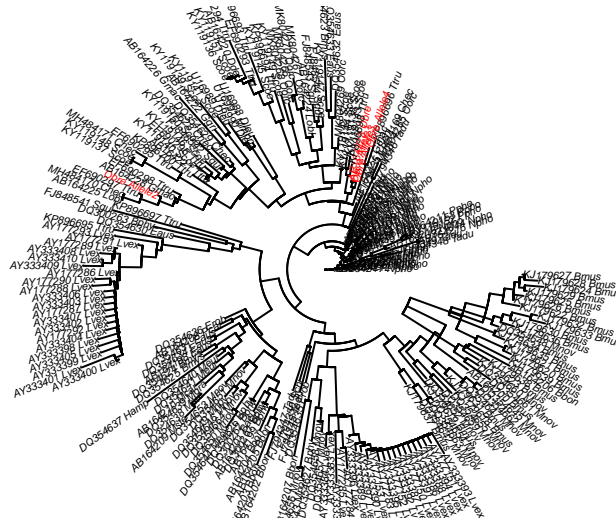
## Using ape package

```
# Load the 'ape' package
library(ape)
tree2 <- read.tree("../Data/output.contree")
# Ladderizing
tree2 <- ladderize(tree2, right = FALSE)

# Define the labels you want to highlight
highlight_labels <- c("Obre_Allele1", "Obre_Allele2", "Obre_Allele3", "Obre_Allele4", "AB164223_Obre")

# Plot the tree with highlighted labels
plot(tree, tip.color = ifelse(tree$tip.label %in% highlight_labels, "red", "black"), cex = 0.3,
      type = 'fan')
```





## References

- Bezuijen, M.R., Zanre, R., Goichot, M. 2007. The Don Sahong Dam and the Irrawaddy dolphin. WWF Greater Mekong Programme, Vientiane
- Beasley, I., Chooruk, S., and Piwpong, N. 2002. THE STATUS OF THE IRRAWADDY DOLPHIN, OR CAELLA BREVIROSTRIS, IN SONGKHLA LAKE, SOUTHERN THAILAND. THE RAFFLES BULLETIN OF ZOOLOGY: Supplement No. 10: 75-83
- Brownell, Jr. R. L., Reeves, R. R., Read A. J., Smith, B. D., Thomas, P. O., Ralls, K., Amano, M., Berggren, P., Chit, A. M., Collins, T., Currey, R., Dolar, M. L. L., Genov, T., Hobbs, R. C., Krebs, D., Marsh, H., Mei Z., Perrin, W. F., Phay, S. ... Wang, J. Y. 2019. Bycatch in gillnet fisheries threatens Critically Endangered small cetaceans and other aquatic megafauna. *Endangered Species Research* 40, 285-296. <https://doi.org/10.3354/esr00994>
- Brownell Jr. R.L., Reeves RR, Thomas PO, Smith BD, Ryan GE. Dams threaten rare Mekong dolphins. *Science*. 2017 Feb 24;355(6327):805. doi: 10.1126/science.aam6406. Epub 2017 Feb 23. PMID: 28232545.
- Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120. (doi: 10.1093/bioinformatics/btu170)
- Budi, T., Piyapattanakorn, S., Krebs, D., Yuda, P., Ninwat, S., Hardwises, P., Prachamkhai, P., Senanan, W., Thongsukdee, S., Phavaphutanon, J., and Klinsawat, W. 2022. Mitogenomes provide insight into complex evolutionary history of freshwater and coastal Irrawaddy dolphin *Orcaella brevirostris* Gray, 1966) in Thailand and Indonesia. *Agriculture and Natural Resources* 56, 591-606. <https://doi.org/10.34044/j.anres.2022.56.3.15>

- Caballero, S., Heimeier, D., Trujillo, F., Vianna, J.A., B-Garrido., Montiel, M.J., Pedereros, S. V., Marmon-  
tel, M., De. O. Santos, M.C., Santos, F.R., and Baker, C.S. 2010. Initial description of Major Histocom-  
patibility Complex variation at two class II loci (DQA-DQB) in *Sotalia fluviatilis* and *Sotalia guianensis*.  
LAJAM 8(1-2): 81-95. e-ISSN 2236-1057
- Cerretero, S.A., Kapli, P., and Yang, Z. 2023. Beginner's Guide on the Use of PAML to Detect Positive  
Selection. Mol. Biol. Evol. 40(4):msad041. <https://doi.org/10.1093/molbev/msad041>
- Dabney J & Meyer M. 2019. Extraction of highly degraded DNA from ancient bones and teeth. In B.  
Shapiro, A. Barlow, P. Heintzman, M. Hofreiter, J. Paijmans, & A. Soares (Eds.), Ancient DNA, Methods  
in Molecular Biology vol. 1963, pp. 25-29. Humana Press. [https://doi.org/10.1007/978-1-4939-9176-1\\_4](https://doi.org/10.1007/978-1-4939-9176-1_4)
- Dai, Y., Chantira, R., Kittiwattana Wong, K., Zhao, L., Sakornwimon, W., Aierken. R., Wu, F., and Wang, X.  
2021. Genetic structure of the endangered Irrawaddy dolphin (*Orcaella brevirostris*) in the Gulf of Thailand.  
Genetics and Molecular Biology 44, e20200365. <https://doi.org/10.1590/1678-4685-GMB-2020-0365>
- Frankham, R. 1995. Conservation Genetics. Annu. Rev. Genet. 29(1):305–327 IUCN–SSC Cetacean  
Specialist Group, “Mekong River Irrawaddy dolphins.”. 2021. <https://iucn-csg.org/mekong-dolphins/>
- Hamner, R.M., Pichler, F.B., Heimeier, D., Constantine, R., and Scott, B. C. 2012. Genetic differentiation  
and limited gene flow among fragmented populations of New Zealand endemic Hector's and Maui's dolphins.  
Conserv Genet 13, 987–1002. <https://doi.org/10.1007/s10592-012-0347-9>
- Hayashi, K., Nishida, S., Yoshida, H., Goto, M., Pastene, L.A., and Koike, H. 2003. Sequence variation of  
the DQB allele in the cetacean MHC. Mammal Study 28: 89–96. DOI: 10.3106/mammalstudy.28.89
- Hughes, A.L., and Hughes, M.K. 1995. Natural Selection of the peptide binding regions of major histocom-  
patibility complex molecules. Immunogenetics 42: 233-243.
- Jackson-Ricketts, J., Ruiz-Cooley, R. I., Junchompoo, C., Thongsukdee, S., Intongkham, A., Ninwat, S.,  
Kittiwattana Wong, K., Hines, E. M., & Costa, D. P. (2019). Ontogenetic variation in diet and habitat of  
Irrawaddy dolphins (*Orcaella brevirostris*) in the Gulf of Thailand and Andaman Sea. Marine Mammal  
Science 35, 492-521. <https://doi.org/10.1111/mms.12547>
- Miller, M.A., Pfeiffer, W., and Schwartz, T. 2011. The CIPRES science gateway: a community resource  
for phylogenetic analyses. Association for Computing Machinery, New York, NY, USA, Article 41, 1-8.  
<https://doi.org/10.1145/2016741.2016785>
- Minton G, Smith BD, Braulik GT, Kreb D, Sutaria D and Reeves R. 2017. *Orcaella brevi-*  
*rostris* (errata version published in 2018), The IUCN Red List of Threatened Species, [https://](https://www.iucnredlist.org/species/15419/12379085)  
[www.iucnredlist.org/species/15419/12379085](https://www.iucnredlist.org/species/15419/12379085)
- Moreno-Santillán, D.D., Lacey, E.A., Gendron,, D., and Ortega, J. 2016. Genetic Variation at Exon of the  
MHC Class II DQB Locus in Blue Whale (*Balaenoptera musculus*) from the Gulf of California. PLoS ONE  
11(1): e0141296. doi:10.1371/journal. pone.0141296
- Murphy, P.T., Irvine, K.N., Sampson, M., Guo, J., and Parr T. 2007. Mercury contamination along the  
Mekong River, Cambodia. Asian Journal of Water, Environment and Pollution. 6(1): 1–9
- Nei, M., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsyn-  
onymous nucleotide substitutions. Molecular Biology and Evolution 3:418-426.
- Schnitzler, J.G., Das, K., Wohlsein, P., Kuiken, T., Ludwig, A., Lieckfeldt, D., Phan, C., Phay, S., and  
Siebert, S. 2021. The Irrawaddy dolphin (*Orcaella brevirostris*) from the Mekong river Cambodia: Prelimi-  
nary health and toxicological investigations. Aquatic Toxicol., 234. <https://doi.org/10.1016/j.aquatox.2021.105812>
- Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conserva-  
tion. Front Zool. doi: 10.1186/1742-9994-2-16 PMID: 16242022
- Thomas, P. O., Gulland, F. M. D., Reeves, R. R., Kreb, D., Ding, W., Smith, B., Malik, M. I., Ryan, G. E.,  
& Phay, S. 2019. Electrofishing as a potential threat to freshwater cetaceans. Endangered Species Research  
39, 207-220. <https://doi.org/10.3354/esr00962>

Tamura, K., Stecher, G., and Kumar, S. 2021. MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* <https://doi.org/10.1093/molbev/msab120>.

Stecher, G., Tamura K., and Kumar, S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* 37:1237-1239.