MHCIIDQB Phylogenetic Tree of the Irrawaddy dolphin in Relation to Cetaceans

Sett Aung

2024-05-10

```
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:

Aim 1: Investigate to potentially uncover additional DQB alleles with larger sample size from populations across its habitat range.

Rationale: The findings of additional DQB alleles could aid in determining the Irrawaddy dolphin population health. The health of the population is related to the variability of the MHC Class II DQB Allele.

Methods: 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 heinsohni (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 sequences from GenBank and align them to the DQB alleles from Irrawaddy dolphin and construct a phylogenetic tree using R and R studio. 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.

```r
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
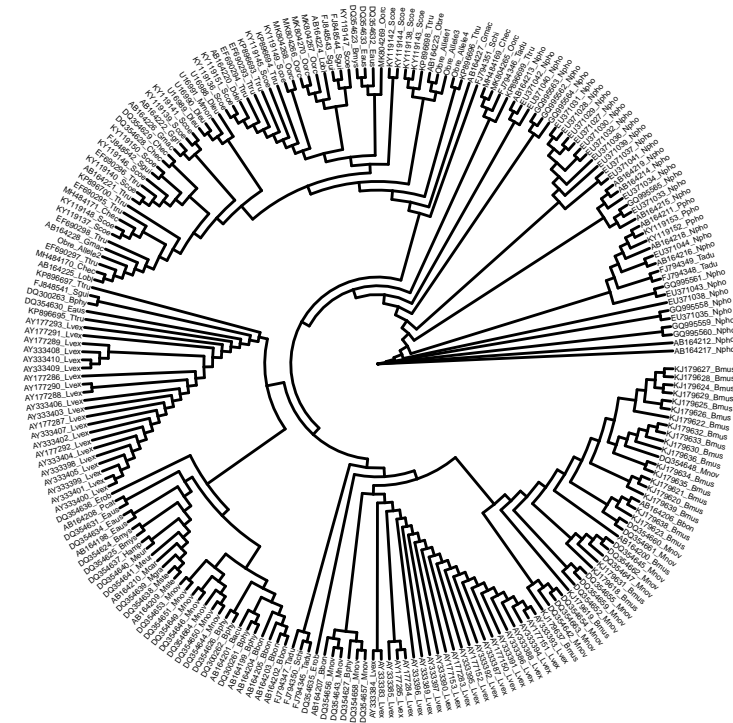


```r
# reading contree output from CIPRES
tree2 <- read.tree("../Data/output.contree")
```

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

```r
# Importing contree from CIPRES using IQ Tool
tree_3 <- 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_3
```

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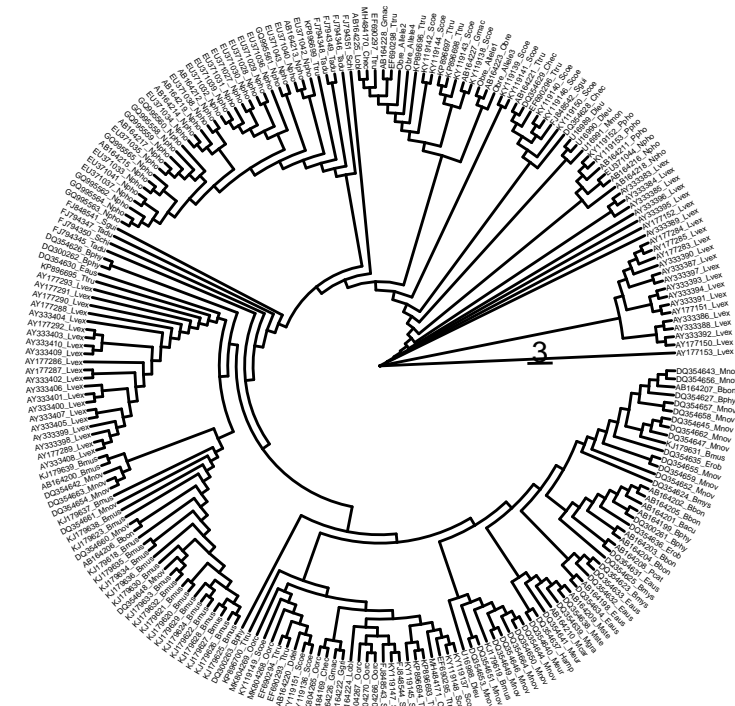

```r
# reading ML treefile output from CIPRES
tree4 <- read.tree("../Data/output.treefile")


# Ladderizing
tree4 <- ladderize(tree4, right = F)


# Importing ML tree from CIPRES using IQ Tool
tree_5 <- ggtree(tree4, 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_5
```

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