Barn Owl Gene Expression – BIOL 710

Pinkal Tandel

Introduction

Malaria is the most common parasitic disease in humans, transmitted by the bite of anopheline mosquitoes carrying Plasmodium spp. parasites [1]. It affects hundreds of millions globally, with an estimated 263 million cases in 2023, resulting in 597,000 deaths across 83 countries [2]. Plasmodium parasites infect red blood cells and belong to the order Haemosporidia, which also includes avian haemosporidians such as Leucocytozoon species. These parasites cause malaria-like symptoms in birds and have been associated with the decline of certain bird species [3]. Leucocytozoon infections are known to induce significant immune responses and physiological stress in birds.

While extensive research has been conducted on haemosporidian parasites, the role of viruses within these protozoa has remained understudied. Avian haemosporidian species show variability in their effects on hosts, with some infections being nearly undetectable, while others cause severe anemia and death. Avian malaria caused by Leucocytozoon parasites is a widespread disease in birds, often leading to significant physiological stress and immune responses in infected hosts.

In 2019, a new class of RNA viruses infecting parasitic protozoans was identified, characterized by a bi-segmented genome consisting of an RNA-dependent RNA polymerase (RdRp) and a hypothetical protein with two overlapping open reading frames (ORFs) of unknown function. The concept of "Matryoshka" infections, where parasites themselves contain viral infections, has been increasingly studied in protozoan biology [4]. These viruses, named Matryoshka RNA viruses (MaRNAV), were first discovered in Plasmodium vivax-infected humans and mosquitoes in eastern Malaysia, with the initial strain designated as MaRNAV-1. Subsequent studies using meta-transcriptomic analysis identified MaRNAV-2 in Leucocytozoon-infected Australian birds [5,6]. These findings suggest that viruses may influence the biology of avian Leucocytozoon parasites, raising important questions about their impact on host-pathogen interactions and disease outcomes.

Further research revealed additional MaRNAV strains in birds infected with haemosporidians: MaRNAV-3 and MaRNAV-6 were detected in American birds carrying Leucocytozoon species, while MaRNAV-4 and MaRNAV-5 were associated with Haemoproteus infections [6]. These findings highlight a broader parasite-virus relationship within avian malaria systems and emphasize the need to investigate host immune responses in co-infection contexts.

Co-infection is increasingly recognized as a key factor in shaping protozoan infections, with evidence suggesting that viral infections in parasites can modulate virulence, immune evasion, and transmission dynamics. Understanding host-pathogen interactions could provide insight into avian immune responses, and understanding how immune genes especially innate immune genes that serve as the initial defense mechanism have evolved in birds, a highly diverse group of tetrapods, is crucial [7].

Co-infections like these may influence parasite behavior and host immune response, but the biological significance of MaRNAVs is still unknown. In avian malaria systems, understanding these interactions could provide valuable insights into immune evolution and disease outcomes.

Barn owls (Tyto alba) were chosen for this study as a model species, as they have been previously utilized in research to assess immune responses in birds infected with Leucocytozoon and MaRNAV. This choice allows for control over species-specific factors, providing a standardized approach to studying gene expression in response to haemosporidian infection. This project builds upon a prior RNA-seq study that identified

immune-related genes differentially expressed in infected owls. That study, however, was limited to only six individuals. In an earlier study by Caroline Faircloth, RNA-seq analysis of barn owls (Tyto alba) identified genes that were differentially expressed in owls infected with both Leucocytozoon and MaRNAV.To build on this foundational work, this project focuses on expanding the dataset by adding more Leucocytozoon-infected barn owls and analyzing expression of a small number of key immune genes for validation.

To strengthen those findings, this study adds new Leucocytozoon-infected barn owl replicates to the dataset and focuses on validating a small number of immune-related genes such as, MAP4K4, HEBP1 previously identified. The ultimate goal is to prepare these targets for downstream ddPCR validation and support the completion of a publishable manuscript.

Research Question

Which immune-related genes are upregulated in barn owls infected with Leucocytozoon, and can we strengthen RNA-seq results through replication and ddPCR validation?

Aims

Aim 1: Expand the Dataset Increase the number of barn owl samples infected with Leucocytozoon, building on the original RNA-seq study. This improves statistical power and robustness of observed gene expression differences.

Aim 2: Focus on Key Immune Genes Analyze expression levels of MAP4K4 and use heatmap visualization to observe global expression clustering by infection status.

Sample Metadata

Sample ID	Condition	Age	Notes
LWC142	Positive	Juvenile	Minor foot wounds; eye irritation
LWC143	Positive	Adult	Minor body wounds; eye irritation
LWC153	Positive	Juvenile	Fell from nest
LWC148	Negative	Juvenile	Euthanized (fracture)
LWC157	Negative	Juvenile	Found in dumpster
LWC162	Negative	Juvenile	Found on ground

- "Positive" = *Leucocytozoon*-infected
- "Negative" = uninfected with both parasite and virus (per Caroline's RT-PCR and RNA-seq screening)

Repository Structure

- `data/`: Contains gene expression CSV and DatasetCard.md
- `owl_gene_analysis.Rmd`: Main R Markdown file with analysis and interpretation
- `README.md`: Project overview

Notes

This work supports a follow-up ddPCR experiment to quantify expression levels of key immune genes. Genes were selected based on differential expression in Caroline's paper and relevance to *Leucocytozoon* or *Plasmodium* infection studies in birds.

Selected Aim for Statistical Analysis

Aim 2: Focus on Key Immune Genes

Specifically: Analyze expression levels of MAP4K4 and use heatmap visualization to observe global expression clustering by infection status.

Hypotheses for MAP4K4 Expression Analysis

- Null Hypothesis: There is no difference in MAP4K4 gene expression between Leucocytozoon-infected and uninfected barn owls.
- Alternative Hypothesis: MAP4K4 gene expression differs between Leucocytozoon-infected and uninfected barn owls.

Statistical Approach & Visualization Description

- To test whether MAP4K4 expression levels differ by infection status, I used a Wilcoxon rank-sum test, a non-parametric alternative to the t-test that is well-suited for small sample sizes and does not assume a normal distribution.
- I visualized expression with a boxplot grouped by infection status (Positive vs. Negative) and overlaid individual sample points to show spread and outliers. I used a viridis color palette, which is colorblind-friendly.
- In addition, I created a heatmap of the top 75 most variable genes, standardized using z-scores across rows, and hierarchically clustered both genes and samples. This allows visual identification of gene clusters and expression trends related to infection status.

Load Libraries

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tibble)
library(tidyr)
library(pheatmap)
library(RColorBrewer)
library(viridis)
```

```
## Loading required package: viridisLite
```

```
# Load merged gene expression data from 6 barn owls
gene_data <- read.csv("data/gene_expression_data.csv")

# Preview the first few rows
head(gene_data)
```

```
##   GeneID LWC142 LWC143 LWC148 LWC153 LWC157 LWC162
## 1   A1CF      0     13      0      0      0      0
## 2 A4GALT     20     82     30     22    149     56
## 3  A4GNT      0      1      0      0      0      0
## 4   AAAS     94    280     53    554    651    199
## 5   AACS     22     16      5     19     12     26
## 6  AADAC      0      9      0      0      7      0
```

Summary Statistics

```
# Number of genes and samples
nrow(gene_data)   # Total number of genes
```

```
## [1] 18892
```

```
ncol(gene_data)   # Number of samples + GeneID column
```

```
## [1] 7
```

```
# Check for missing values
sum(is.na(gene_data))   # Should be 0
```

```
## [1] 0
```

```
# View expression counts for MAP4K4
gene_data[gene_data$GeneID == "MAP4K4", ]
```

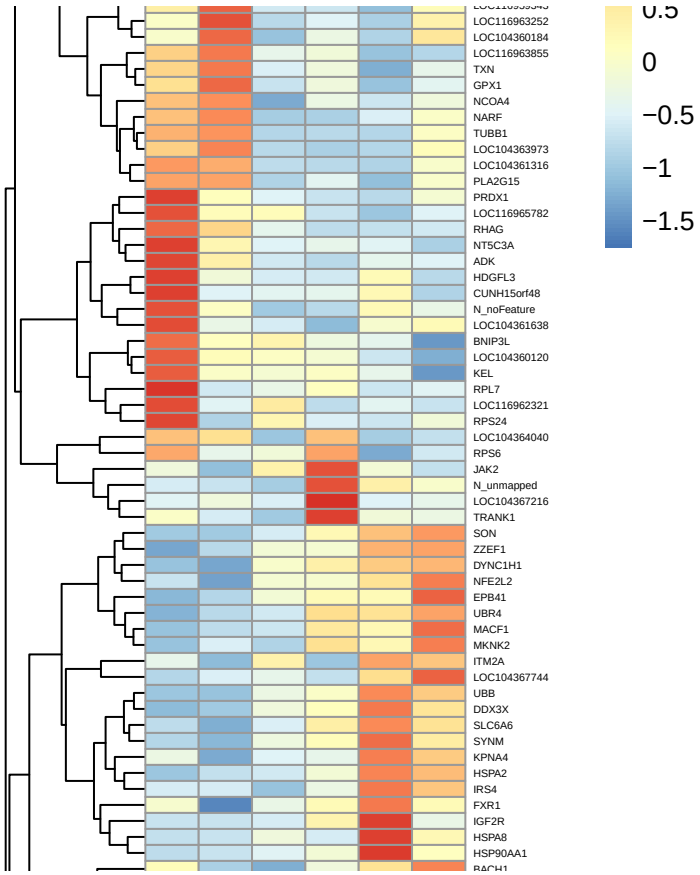
```
##       GeneID LWC142 LWC143 LWC148 LWC153 LWC157 LWC162
## 12062 MAP4K4    114    155    285    134    757    377
```

Heatmap: Top 75 Most Variable Genes

```
gene_matrix <- gene_data %>% column_to_rownames("GeneID") %>% as.matrix()
var_genes <- apply(gene_matrix, 1, var)
top_genes <- sort(var_genes, decreasing = TRUE)[1:75]
filtered_matrix <- gene_matrix[names(top_genes), ]
z_matrix <- t(scale(t(filtered_matrix)))
```

```r
annotation_col <- data.frame(
  Condition = c("Positive", "Positive", "Negative", "Positive", "Negative", "Negative")
)
rownames(annotation_col) <- colnames(z_matrix)

pheatmap(
  z_matrix,
  annotation_col = annotation_col,
  color = colorRampPalette(rev(brewer.pal(n = 7, name = "RdYlBu")))(100),
  clustering_distance_rows = "euclidean",
  clustering_distance_cols = "euclidean",
  clustering_method = "complete",
  fontsize_row = 4,
  fontsize_col = 12,
  cellheight = 6,
  cellwidth = 20,
  angle_col = 45,
  main = "Top 75 Most Variable Genes across 6 Barn Owls"
)
```

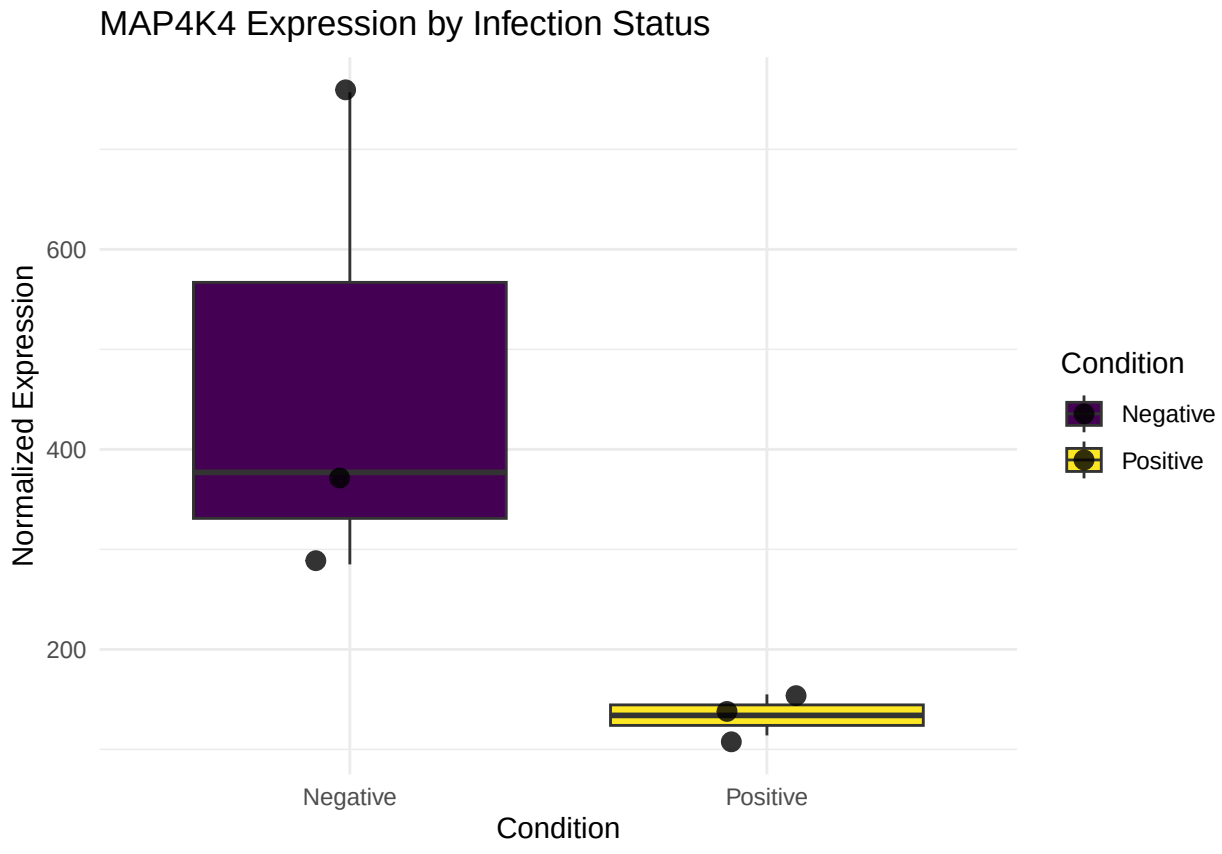


MAP4K4 Expression Analysis

```r
sample_info <- data.frame(
  Sample = colnames(gene_data)[-1],
  Condition = c("Positive", "Positive", "Negative", "Positive", "Negative", "Negative")
)

map4k4 <- gene_data %>%
  filter(GeneID == "MAP4K4") %>%
  pivot_longer(-GeneID, names_to = "Sample", values_to = "Expression") %>%
  left_join(sample_info, by = "Sample")

# Plot
# This is GOOD (directly plotted)
ggplot(map4k4, aes(x = Condition, y = Expression, fill = Condition)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(width = 0.1, size = 3, alpha = 0.8) +
  scale_fill_viridis_d(option = "D") +
  theme_minimal() +  # This is fine
  labs(
    title = "MAP4K4 Expression by Infection Status",
    y = "Normalized Expression",
    x = "Condition"
  )
```



Wilcoxon Test

```
wilcox.test(Expression ~ Condition, data = map4k4)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  Expression by Condition
## W = 9, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
```

Results

- The boxplot shows that MAP4K4 expression appears higher in uninfected owls compared to infected ones, suggesting a potential immune-related downregulation in response to infection.
- The Wilcoxon test yielded a p-value of 0.1, which is above the typical significance of 0.05, meaning we do not reject the null hypothesis. However, the small sample size limits statistical power. This is why aim 1 works towards increasing replicates.
- The heatmap reveals clear clustering of gene expression patterns by infection status, supporting the idea that infection influences broader immune gene expression, even if individual gene tests lack statistical power, which could be increased by following aim1 and adding more replicates.

References

1. Ferraboli JW, Veiga GTS da, Albrecht L. Plasmodium vivax transcriptomics: What is new? Exp Biol Med. 2023;248(19):1645. doi:10.1177/15353702231198070
2. Fact sheet about malaria. Accessed March 7, 2025. https://www.who.int/news-room/fact-sheets/detail/malaria
3. Valkiūnas G. Avian Malaria Parasites and Other Haemosporidia. CRC Press; 2005.
4. Rodrigues JR, Roy SW, Sehgal RNM. Novel RNA viruses associated with avian haemosporidian parasites. PloS One. 2022;17(6):e0269881. doi:10.1371/journal.pone.0269881
5. Petrone ME, Charon J, Grigg MJ, et al. A virus associated with the zoonotic pathogen Plasmodium knowlesi causing human malaria is a member of a diverse and unclassified viral taxon. Published online September 19, 2024. doi:10.1101/2024.09.18.613759
6. Duc M, Esperanza C, Chagas CRF, Iezhova T, Sehgal RNM, Valkiūnas G. First report of Matryoshka RNA virus in an African-European migrant bird. PLOS ONE. 2025;20(3):e0319395. doi:10.1371/journal.pone.0319395
7. Outlaw DC, Walstrom VW, Bodden HN, Hsu C Yu, Arick M, Peterson DG. Molecular evolution in immune genes across the avian tree of life. Parasitol Open. 2019;5:e3363. doi:10.1017/pao.2019.33