Independent Project Introduction

Timara Vereen

2025-03-05

Introduction

Vector behavior is a key factor in disease transmission, influencing pathogen spread among humans, wildlife, and domestic animals. Various arthropod vectors, such as mosquitoes, fleas, and tsetse flies, exhibit host-seeking behaviors that shape disease dynamics (Benelli, 2020; Onen et al., 2023; Rogers & Bates, 2007; Pipová et al., 2023). Vector-borne diseases account for over 17% of infectious diseases, leading to more than 700,000 deaths annually (World Health Organization, 2020; Sutherust, 2004; Chala & Hamde, 2021). Climate change, land-use alterations, and human activity are driving the increasing prevalence of these diseases by disrupting ecological systems and increasing host-vector interactions (Ellwanger et al., 2020; Swei et al., 2020). While the behaviors of many vectors have been documented (Azad et al., 1997; Gleave et al., 2016; Bruno et al., 2019; Javed et al., 2021; Stanczyk et al., 2017; Wei Xiang et al.), nymphal tick questing behavior, particularly in Ixodes pacificus, remains less understood, despite its implications for disease transmission.

Ixodes pacificus, the primary vector of Lyme disease (LD) in the western U.S. (CDC, 2020), plays a crucial role in pathogen transmission, with over 400,000 new cases estimated annually in the U.S. (Garira & Chirove, 2019; CDC, 2024). While Ixodes scapularis nymphs in the eastern U.S. exhibit well-documented host-seeking behaviors, I. pacificus questing activity is more variable and less frequently observed, particularly in Southern California (Lane & Kjemtrup, 2016; Lane et al., 2013). Geographic, genetic, and environmental factors likely contribute to these behavioral differences, influencing LD incidence across regions (Arsnoe et al., 2019). Additionally, host interactions, such as those with Western fence lizards (Sceloporus occidentalis), may shape tick activity, as lizard-associated tick abundances fluctuate seasonally (Lane & Loye, 1989). These variations in questing behavior highlight the need for further research to understand the environmental and ecological drivers of tick activity, which could aid in disease prevention and risk assessment (Tomkins et al., 2014). Our goals are to: 1) Does geographic origin affect I. pacificus questing behavior? I hypothesize that questing behavior will differ based on geographic origin during observation trials. Specific Aim 1: Determine whether geographic origin affects questing behavior due to different environmental conditions in these regions

```r
# Load dataset -test
#test #2
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.3
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```
## Warning: package 'dplyr' was built under R version 4.4.2
```

```
## -- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
```

```
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
behavior_data <- read.csv("Master's data - results_4.1.25.csv")
behavior_data$Engorgement <- as.factor(behavior_data$Engorgement)

# View first few rows
view(behavior_data)
```

Aim 1

Specific Aim 1: Determine whether geographic origin affects questing behavior due to different environmental conditions in these regions

Hypotheses

Null Hypothesis: Geographic origin has no effect on tick questing behavior, regardless of the environmental conditions associated with different regions.

Alternative Hypothesis: Geographic origin affects tick questing behavior, likely due to differences in environmental conditions between regions.

Statistical approach

Behavioral trials were conducted in two different laboratory rooms, each with distinct environmental conditions. The first room, maintained at approximately 17.46°C and 70.18% relative humidity (hereafter referred to as the "17.46°C room"), was located in the San Francisco Animal Care Facility and housed four mus musculus (C3H) mice as part of another experiment. Initial trials were conducted in this space, which offered more consistent temperature and humidity control. For this project, I will be reporting on data from the 17.46°C room.

All statistical analyses were conducted in R (version 4.4.1) using RStudio, with key packages including tidyverse, glmmTMB, performance, and ggplot2 for data cleaning, visualization, and modeling. Questing behavior metrics were analyzed using Generalized Linear Mixed Models (GLMMs) to account for non-normal distributions and repeated measures. Total distance, horizontal distance, and vertical gross distance were modeled using GLMMs with a Tweedie distribution to account for overdispersion and the continuous nature of the distance data. Percent activity, a proportion-based metric with many zero values, was analyzed using a negative binomial GLMM with a zero-inflated component. For the sake of this project, I will be reporting on the total distance traveled behavioral metric. Fixed effects included tick population of origin and engorgement level across all models. Tick ID and trial number were included as random effects to account for repeated observations and individual variation. Model performance was evaluated using Akaike Information Criterion (AIC), with lower AIC values guiding model selection.

Stats and Visuals

```r
#install these packages if you haven't :)
#install.packages("ggplot2")
#install.packages("ggsignif")
library(ggplot2)
library(ggsignif)
```

```
## Warning: package 'ggsignif' was built under R version 4.4.3
```

```r
library(glmmTMB)
```

```
## Warning: package 'glmmTMB' was built under R version 4.4.3
```

```
## Warning in check_dep_version(dep_pkg = "TMB"): package version mismatch:
## glmmTMB was built with TMB package version 1.9.17
## Current TMB package version is 1.9.15
## Please re-install glmmTMB from source or restore original 'TMB' package (see '?reinstalling' for more
```

```r
library(emmeans)
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```r
behavior_data$horizontal_dist_cm <- behavior_data$horizontal_dist * 0.635
behavior_data$vertical_dist_gross_cm <- behavior_data$vertical_dist_gross * 0.635
behavior_data$total_dist_cm <- behavior_data$total_dist * 0.635
behavior_data_u <- subset(behavior_data, Room == "U") # U refers to 17.46 room
```

```r
#total distance (u)
model_total<- glmmTMB(
  total_dist_cm ~ Population * Engorgement + (1|Tick_ID) + (1|Trial),
  family = tweedie(link = "log"),
  data = behavior_data_u
)

summary(model_total)
```

```
##  Family: tweedie  ( log )
## Formula:
## total_dist_cm ~ Population * Engorgement + (1 | Tick_ID) + (1 |     Trial)
## Data: behavior_data_u
##
##      AIC      BIC   logLik deviance df.resid
##   1366.6   1400.8   -673.3   1346.6      215
##
## Random effects:
##
## Conditional model:
##  Groups  Name        Variance  Std.Dev.
```

```
##  Tick_ID (Intercept) 7.547e-01 8.687e-01
##  Trial   (Intercept) 3.020e-10 1.738e-05
## Number of obs: 225, groups:  Tick_ID, 75; Trial, 3
##
## Dispersion parameter for tweedie family (): 1.76
##
## Conditional model:
##                                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)                              2.1851     0.2001  10.918  < 2e-16
## PopulationSouthern California           -1.7981     0.6093  -2.951  0.00317
## Engorgement2                            -0.3764     0.3171  -1.187  0.23522
## Engorgement3                            -0.1390     0.3708  -0.375  0.70774
## PopulationSouthern California:Engorgement2  1.1607  0.7361   1.577  0.11484
## PopulationSouthern California:Engorgement3  1.6786  0.7207   2.329  0.01986
##
## (Intercept)                              ***
## PopulationSouthern California            **
## Engorgement2
## Engorgement3
## PopulationSouthern California:Engorgement2
## PopulationSouthern California:Engorgement3 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
model_total1<- glmmTMB(
  total_dist_cm ~ Population + Engorgement + (1|Tick_ID) + (1|Trial),
  family = tweedie(link = "log"),
  data = behavior_data_u
)

summary(model_total1)
```

```
##  Family: tweedie  ( log )
## Formula:
## total_dist_cm ~ Population + Engorgement + (1 | Tick_ID) + (1 |      Trial)
## Data: behavior_data_u
##
##      AIC      BIC   logLik deviance df.resid
##   1368.1   1395.4   -676.0   1352.1      217
##
## Random effects:
##
## Conditional model:
##  Groups  Name         Variance  Std.Dev.
##  Tick_ID (Intercept) 8.238e-01 9.076e-01
##  Trial   (Intercept) 2.776e-10 1.666e-05
## Number of obs: 225, groups:  Tick_ID, 75; Trial, 3
##
## Dispersion parameter for tweedie family (): 1.76
##
## Conditional model:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     2.0423     0.1989  10.266   <2e-16 ***
## PopulationSouthern California   -0.6546     0.2697  -2.428   0.0152 *
```

4

```
## Engorgement2                          -0.2357     0.2904  -0.812    0.4169
## Engorgement3                           0.3285     0.3090   1.063    0.2878
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
AIC(model_total, model_total1)
```

```
##              df      AIC
## model_total  10 1366.598
## model_total1  8 1368.054
```

```r
#model_total has a lower AIC score, so we used that model (model_total)

#conducted an emmeans to see any potential additional interactions
emmeans(model_total, pairwise ~ Population * Engorgement, adjust = "bonferroni")
```

```
## $emmeans
##  Population          Engorgement emmean    SE  df asymp.LCL asymp.UCL
##  Northern California 1            2.185 0.200 Inf     1.793      2.58
##  Southern California 1            0.387 0.579 Inf    -0.749      1.52
##  Northern California 2            1.809 0.248 Inf     1.323      2.29
##  Southern California 2            1.171 0.336 Inf     0.512      1.83
##  Northern California 3            2.046 0.313 Inf     1.433      2.66
##  Southern California 3            1.927 0.246 Inf     1.445      2.41
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
##
## $contrasts
##  contrast                                                          estimate
##  Northern California Engorgement1 - Southern California Engorgement1    1.798
##  Northern California Engorgement1 - Northern California Engorgement2    0.376
##  Northern California Engorgement1 - Southern California Engorgement2    1.014
##  Northern California Engorgement1 - Northern California Engorgement3    0.139
##  Northern California Engorgement1 - Southern California Engorgement3    0.258
##  Southern California Engorgement1 - Northern California Engorgement2   -1.422
##  Southern California Engorgement1 - Southern California Engorgement2   -0.784
##  Southern California Engorgement1 - Northern California Engorgement3   -1.659
##  Southern California Engorgement1 - Southern California Engorgement3   -1.540
##  Northern California Engorgement2 - Southern California Engorgement2    0.637
##  Northern California Engorgement2 - Northern California Engorgement3   -0.237
##  Northern California Engorgement2 - Southern California Engorgement3   -0.118
##  Southern California Engorgement2 - Northern California Engorgement3   -0.875
##  Southern California Engorgement2 - Southern California Engorgement3   -0.755
##  Northern California Engorgement3 - Southern California Engorgement3    0.119
##     SE  df z.ratio p.value
##  0.609 Inf   2.951  0.0475
##  0.317 Inf   1.187  1.0000
##  0.390 Inf   2.600  0.1397
##  0.371 Inf   0.375  1.0000
##  0.314 Inf   0.824  1.0000
##  0.626 Inf  -2.272  0.3464
##  0.664 Inf  -1.182  1.0000
```

```
##  0.657 Inf  -2.526  0.1728
##  0.619 Inf  -2.487  0.1934
##  0.416 Inf   1.532  1.0000
##  0.398 Inf  -0.596  1.0000
##  0.345 Inf  -0.341  1.0000
##  0.459 Inf  -1.908  0.8467
##  0.412 Inf  -1.831  1.0000
##  0.396 Inf   0.301  1.0000
##
## Results are given on the log (not the response) scale.
## P value adjustment: bonferroni method for 15 tests
```

```r
behavior_data_u$Pop_Engorgement <- paste(behavior_data_u$Population, behavior_data_u$Engorgement, sep =

# Okabe-Ito palette for two groups
okabe_ito <- c("Northern California" = "#E69F00",  # Orange
               "Southern California" = "#56B4E9")  # Sky Blue

#total distance traveled by population
ggplot(behavior_data_u, aes(x = as.factor(Room), y = total_dist_cm, fill = Population)) +
  geom_boxplot(outlier.shape = NA, width = 0.6, position = position_dodge(0.75)) +
  stat_summary(fun = mean, geom = "point", shape = 20, size = 3, color = "black",
               position = position_dodge(0.75)) +
  labs(
    x = "Room",
    y = "Total Distance (cm)",
    fill = "Population"
  ) +
  theme_minimal() +
  scale_fill_manual(values = okabe_ito) +
  geom_signif(
    comparisons = list(c("U", "U")),  # You may want to check this comparison
    annotations = "p = 0.032*",
    y_position = max(behavior_data_u$total_dist_cm) * 0.8,
    tip_length = 0.08,
    textsize = 4
  )+
  theme(
  plot.title = element_text(face = "bold", hjust = 0.5, size = 18),
  axis.title = element_text(face = "bold", size = 14),
  axis.text = element_text(size = 12),
  axis.text.x = element_text(angle = 45, hjust = 1),
  legend.title = element_text(face = "bold", size = 13),
  legend.text = element_text(size = 11)
)
```
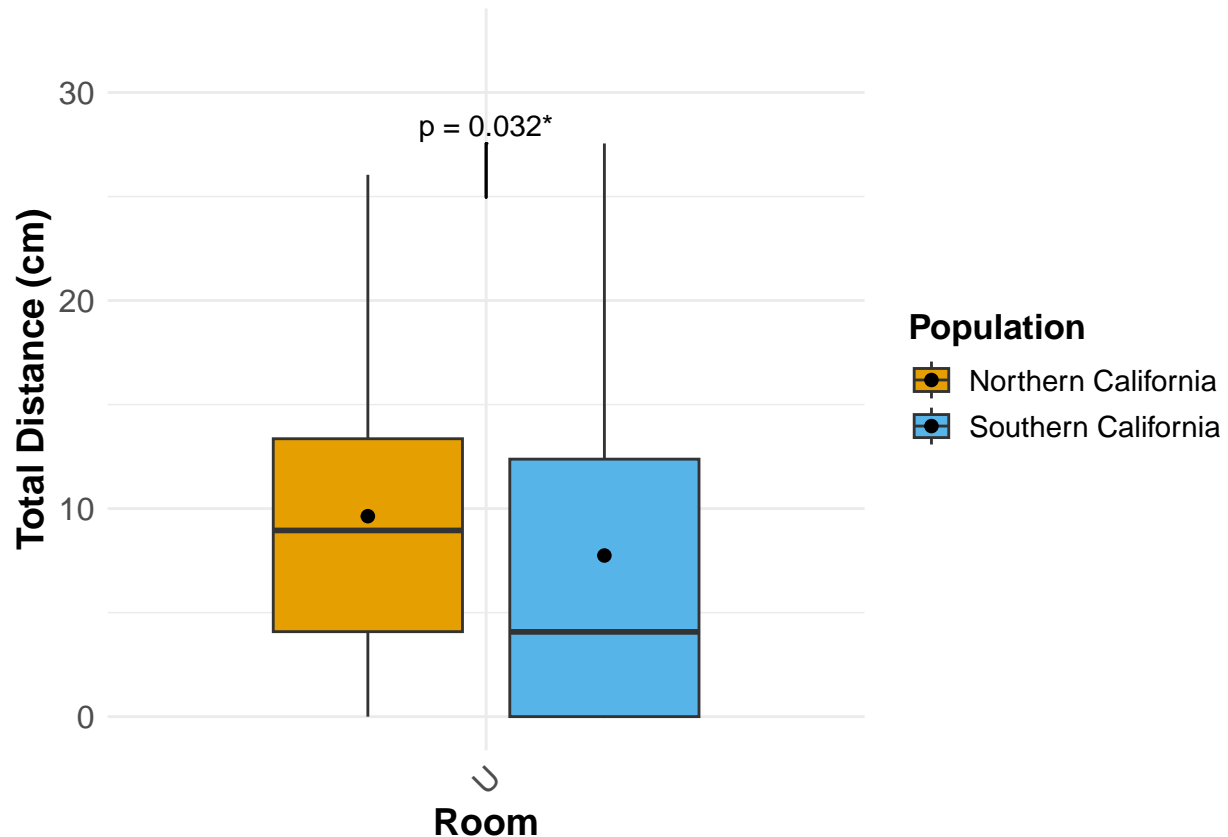
```r
#total distance traveled by population and engorgement
# Okabe-Ito for engorgement levels (three distinguishable colors)
engorgement_colors <- c("1" = "#E69F00",   # Orange (Unfed)
                        "2" = "#56B4E9",   # Sky Blue (Partially Engorged)
                        "3" = "#009E73")   # Bluish-Green (Fully Engorged)

ggplot(behavior_data_u, aes(x = Pop_Engorgement, y = total_dist_cm, fill = as.factor(Engorgement))) +
  geom_boxplot(outlier.shape = NA, width = 0.6) +
  stat_summary(fun = mean, geom = "point", shape = 20, size = 3, color = "black") +
  labs(
    x = "Population",
    y = "Total Distance (cm)",
    fill = "Engorgement Level"
  ) +
  scale_fill_manual(
    values = engorgement_colors,
    labels = c("1" = "Unfed", "2" = "Partially Engorged", "3" = "Fully Engorged")
  ) +
  theme_minimal() +
  geom_signif(
    comparisons = list(
      c("Northern California\n1", "Southern California\n1"),
      c("Southern California\n1", "Southern California\n3")
    ),
    annotations = c("p=0.048", "p=0.020"),
    y_position = c(25, 35),
```
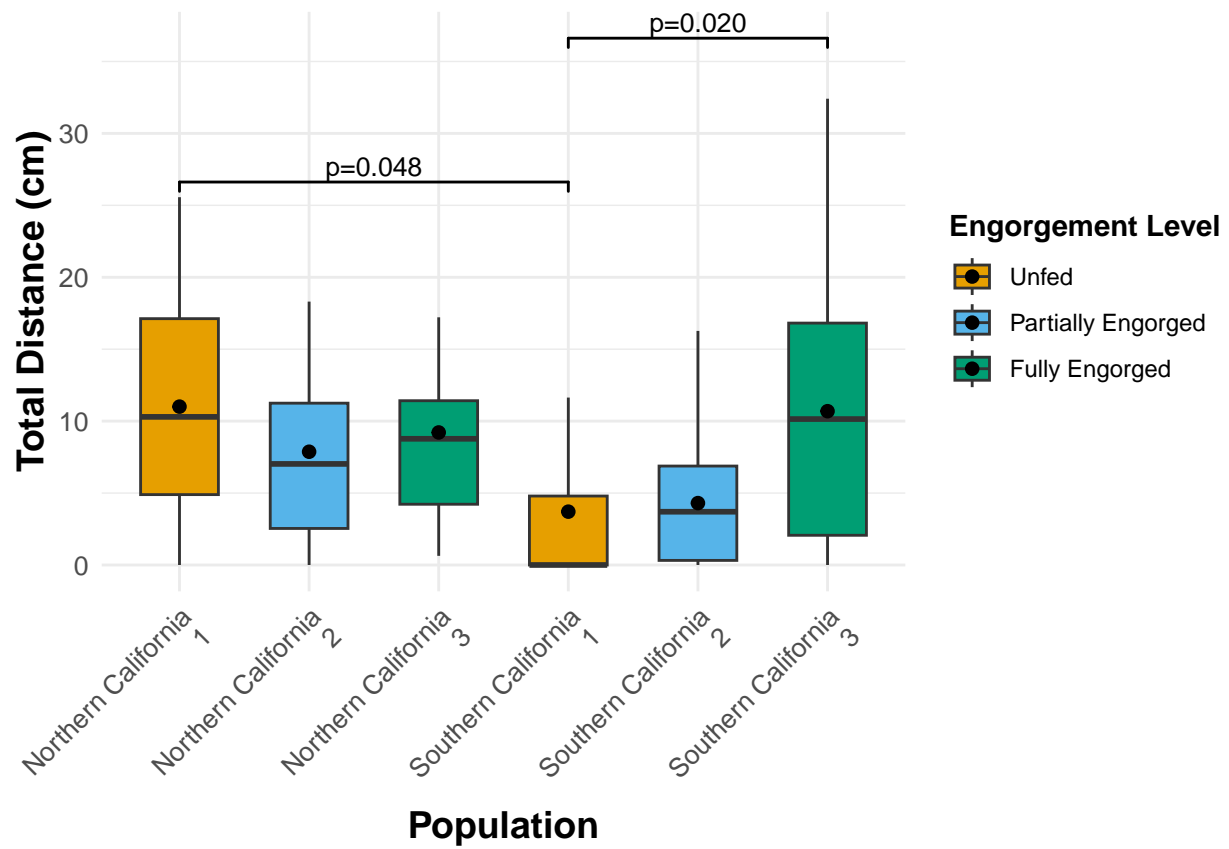
```
    tip_length = 0.02,
    textsize = 3.5,
    map_signif_level = FALSE
) +
theme(
  plot.title = element_text(face = "bold", hjust = 0.5, size = 16),
  axis.title = element_text(face = "bold", size = 14),
  axis.text = element_text(size = 10),
  axis.text.x = element_text(angle = 45, hjust = 1),
  legend.title = element_text(face = "bold"),
  legend.position = "right"
)
```



```
"""
```