

# Report on Figures and Data Analysis

*P. Alexander Burnham*

*January 3, 2017*

## Comparing colony productivity and pathogen infection between California-bred and New England-bred honey bees (*Apis mellifera*)

### Metadata:

**Authors:** Andre J. Burnham, Fiona McLaughlin, P. Alexander Burnham & Herman Lehman

**Date:** 10 August 2016

**Data Set:** These data were collected in 2016 in New York state by Andre Burnham and the Hamilton College bee research team from two yards in NY owned by one beekeeper who volunteered for the study.

**Data Source:** 2016 Hamilton Queen Origin Research Project

**Funding Source:** Hamilton College undergraduate research grants

**Data Collection:** Weights of colonies were obtained by summing the weights of individual hive bodies and supers for each colony at each yard (add more about protocols for other assays).

**Columns:** (from left to right) FieldID: Id including number location and treatment, Time: Generic Time Steps (T1-T5), Origin (California and Local), Yard (1 or 2), Nosema (spores per bee) only three time steps, Varroa (mites per 300 bees), Brood (frames of brood in colony), NosemaPA (binary data for nosema), VarroaPA (binary data for mites), MassDay (days after requeening for mass), NosemaDay (days after requeening for nosema), VarroaDay (days after requeening for varroa), BroodDay (days after requeening for brood), MassDate (dates sampled), NosemaDate (dates sampled), VarroaDate (dates sampled), BroodDate (dates sampled)

**Rows:** Data points for all columns in order from each collection event

**Missing values:** NA

## SUMMARY

### Background

Honey bees (*Apis mellifera*) are global pollinators that have been on the decline due to various environmental and pathogen-related factors (vanEngelsdorp et al. 2009). However, recent studies have suggested that local environmental adaptations in *A. mellifera* phenotypes could be tied to colony productivity, survivability, and other fitness-related factors and thus hinder colony performance when migratory beekeeping practices are exercised (Costa et al. 2012, Buechler et al. 2014). The behavioral and morphological characteristics of these colony ecotypes are largely a result of genotypic variation in drones and the queen bee, a colony's only reproducing female. In this study, we tested the null hypothesis that queen bee ecotypes from very contrasting climates produce offspring that are equally productive and resistant to pathogens when observed in the same geographical region. In particular, we compared queens grafted in Vermont with queens reared by one of the largest queen breeders in California. We conducted our experiment in Central New York, which we determined to be a fair representation of the Vermont bees' local environment.

### Results

We demonstrated that local colonies grew significantly more in total colony biomass than those of the Californian bees during specific time periods (T2, 7 July 2016; and T4, 9 August 2016). We also found

that over five time periods (from 30 May 2016 to 3 August 2016), locally-raised bees reared significantly more brood than the Californian bees. However, no notable correlation with queen bee origin and mass of pollen collected was observed. In assessing resistance to parasites and pathogens, we found no significant difference over time in levels of the ectoparasite *Varroa destructor*. Using semi-quantitative PCR to evaluate the levels of deformed wing virus (DWV), black queen cell virus, (BQCV), and Israeli acute paralysis virus (IAPV) (three common RNA viruses in *A. mellifera*), we found that the levels of viral infection for BQCV and IAPV were unaffected after requeening the two groups with our experimental queens. However, our results demonstrated higher infection levels of DWV in CA colonies than in local colonies from T1 to T2 (15 June 16 and 13 July 2016). Levels of *Nosema* sp. (a microsporidian) in both the local and CA colony groups differed through three time periods. From T1 to T2 (16 June 2016 and 19 July 2016), *Nosema* spore loads significantly decreased in the local bees, while spore loads in the CA colonies increased. The third time period (9 August 2016) also showed significant decreases in local colony infection, while CA colonies continued to have high spore loads.

## Conclusion

The significant differences we observed in colony weight and brood production between California-bred and northern-bred bees suggest that locally-bred *A. mellifera* have stronger nectar foraging and brood-rearing tendencies than their Californian counterparts. The lower *Nosema* sp. spore and DWV loads in the local colonies that we demonstrated strongly suggest that local bees are more resistant to certain pathogen strains and perhaps have heightened immune gene expression when compared with bees from another climate. To our knowledge, this is the first study to demonstrate significant differences in *Nosema* sp. infections between different *A. mellifera* ecotypes. Overall, our results present evidence that offspring of locally-raised queens from the Northeast have better adapted for survival and productivity in northern regions as compared to queens and worker bees from California, a climatically contrasting environment. These environmental adaptations relating to *A. mellifera* fitness could be tied to future colony survivability.

```
# Preliminaries:
```

```
# Clear memory of characters:  
ls()
```

```
## character(0)
```

```
rm(list=ls())
```

```
# Set Working Directory:  
setwd("~/Desktop/QueenExperimentBurnham")
```

```
# Read in Data:  
QueenDF <- read.table("2016QueensHam.csv", header=TRUE, sep = ",")
```

```
# These dataframes are seperate becuae they do not follow that sampling  
# pattern that the other four varaibles in the previous data frame do (i.e.5  
# time steps):
```

```
PollenDF <- read.table("PollenQueens.csv", header=TRUE, sep = ",")  
VirusDF <- read.table("RNAVirus.csv", header=TRUE, sep = ",")
```

```
# load plyr for data manipulation and ggplot plotting package and other  
# related packages:  
library(plyr)  
library(ggplot2)
```

```
library(grid)
library(dplyr)
library(scales)
```

## Colony Mass:

```
# MASS - use repeated measures anova to look at differences between local  
# and california through time:
```

```
aov.out <- aov(Mass~Origin * Time + Error(FieldID), data=QueenDF)
```

```
# look at summary of model: SIGNIFICANT
```

```
summary(aov.out)
```

```
##
## Error: FieldID
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Origin      1   3189     3189   7.840 0.00836 **
## Time        2   2725     1363   3.350 0.04699 *
## Origin:Time  2    124         62   0.153 0.85869
## Residuals   34  13830      407
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Time        3   9513     3171  41.909 < 2e-16 ***
## Origin:Time  3   1125      375   4.958 0.00328 **
## Residuals   81   6129       76
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Frames of Brood:

```
# BROOD - use repeated measures anova to look at differences between local  
# and california through time:
```

```
aov.out <- aov(Brood ~ Origin * Time + Error(FieldID), data=QueenDF)
```

```
# look at summary of model: SIGNIFICANT
```

```
summary(aov.out)
```

```
##
## Error: FieldID
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Origin      1  18.78    18.784  11.748 0.00169 **
## Time        4  77.61    19.403  12.135 4.1e-06 ***
## Origin:Time  2   0.44     0.218   0.136 0.87322
## Residuals   32  51.17     1.599
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Error: Within
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Time           4   63.68   15.920    14.14 1.59e-09 ***
## Origin:Time     4   35.97    8.993     7.99 9.27e-06 ***
## Residuals    123  138.43    1.125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

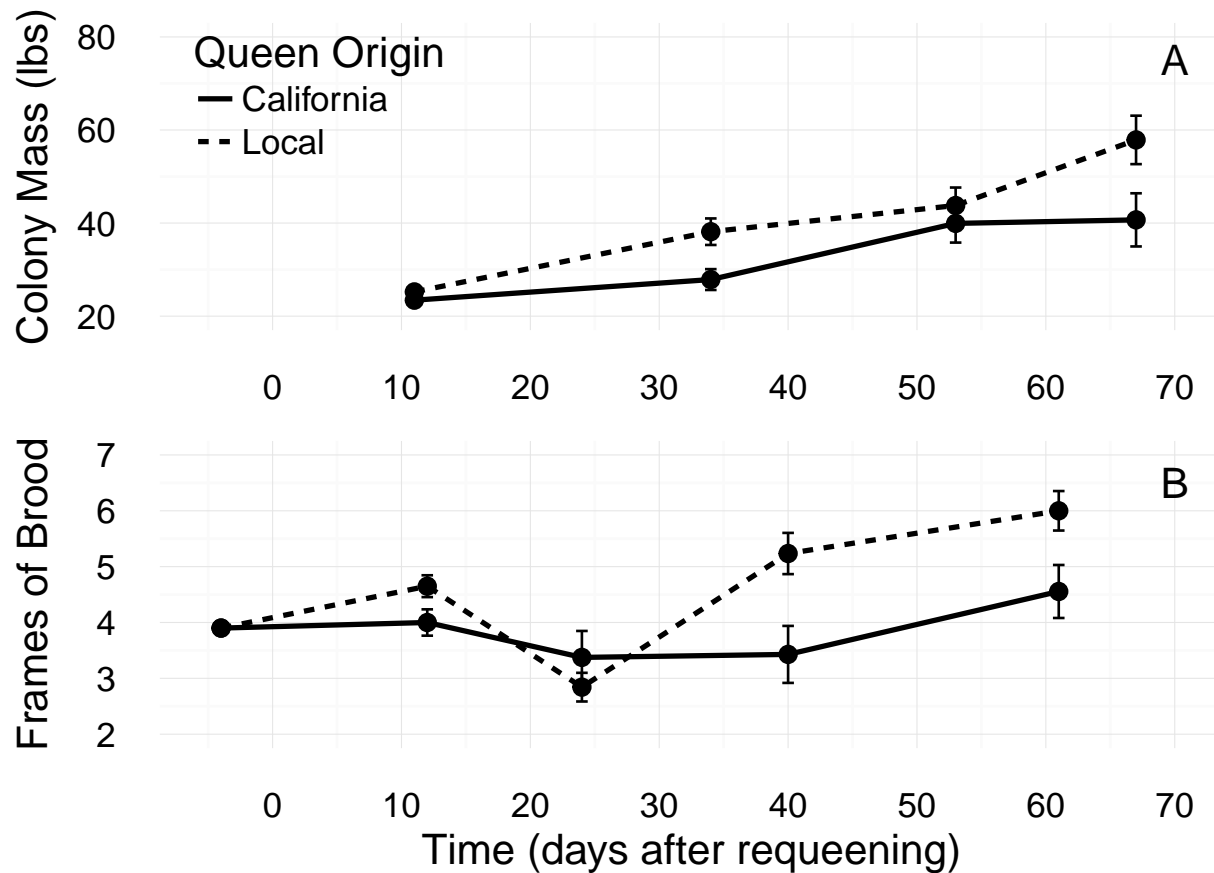
## Frames of Brood & Colony Mass Through Time by Queen Origin

```
##           Origin MassDay  n    mean      sd      se
## 1 California      11 20 23.43000  5.096655 1.139647
## 2 California      34 16 27.86250  8.960571 2.240143
## 3 California      53 14 39.93333 15.427573 4.123192
## 4 California      67  9 40.68889 17.132750 5.710917
## 6      Local      11 20 25.21000  4.309341 0.963598
## 7      Local      34 19 38.14737 12.422711 2.849965
## 8      Local      53 17 43.77647 15.910355 3.858828
## 9      Local      67 17 57.87059 21.473465 5.208080
```

```
##           Origin BroodDay  n    mean      sd      se
## 1 California      -4 20 3.900000 0.3077935 0.06882472
## 2 California      12 20 4.000000 1.0540926 0.23570226
## 3 California      24 16 3.375000 1.8929694 0.47324236
## 4 California      40 14 3.428571 1.9100659 0.51048658
## 5 California      61  9 4.555556 1.4240006 0.47466687
## 6      Local      -4 20 3.900000 0.3077935 0.06882472
## 7      Local      12 20 4.650000 0.8750940 0.19567696
## 8      Local      24 19 2.842105 1.1186876 0.25664454
## 9      Local      40 17 5.235294 1.5218990 0.36911472
## 10     Local      61 17 6.000000 1.4577380 0.35355339
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##      combine
```



Varroa:

```
# VARROA - use repeated measures anova to look at differences between local
# and california through time:
aov.out <- aov(Varroa ~ Origin * Time + Error(FieldID), data=QueenDF)

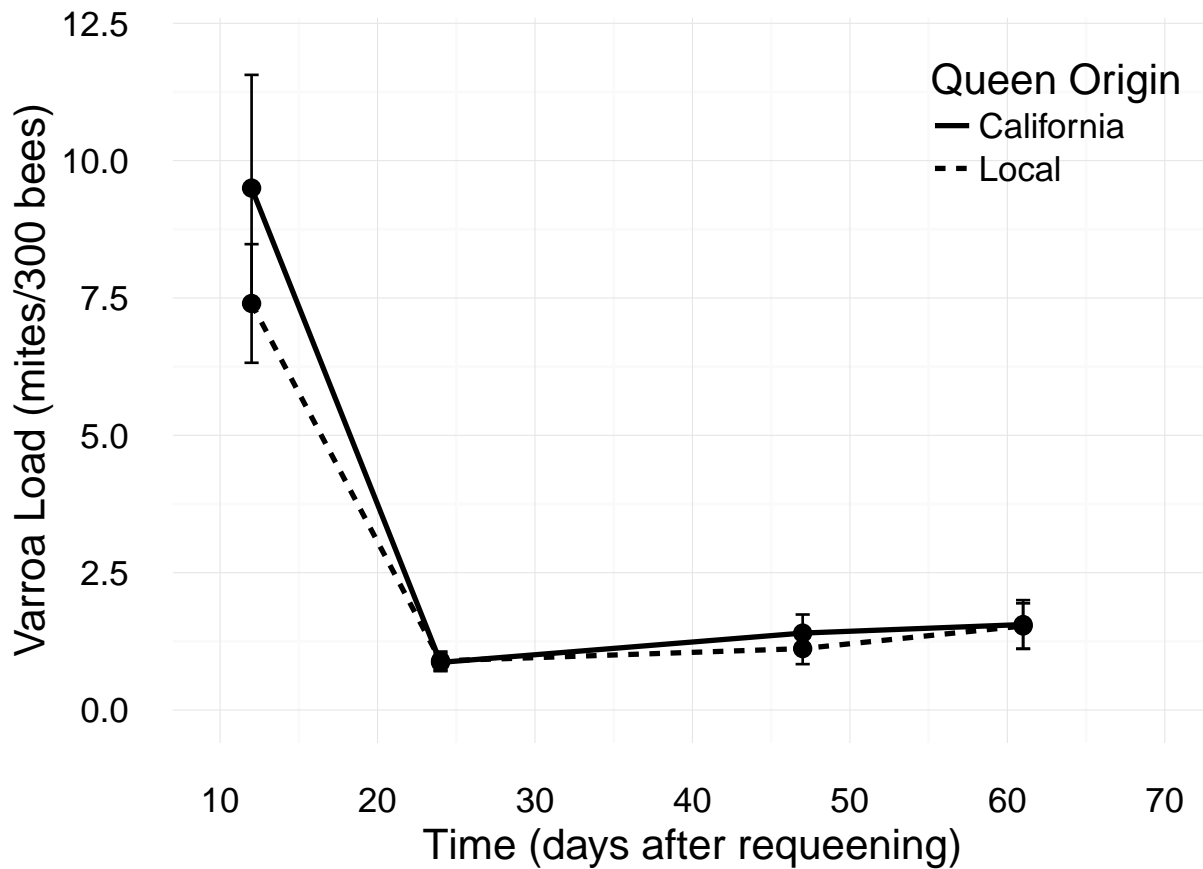
# look at summary of model: NOT SIGNIFICANT
summary(aov.out)
```

```
##
## Error: FieldID
##           Df Sum Sq Mean Sq F value Pr(>F)
## Origin      1  43.7   43.75    3.082  0.089 .
## Time         3   16.0    5.34    0.376  0.771
## Origin:Time  2   26.2   13.10    0.923  0.408
## Residuals   31 439.9   14.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Time         3 1487.8   495.9  30.409 2.87e-13 ***
## Origin:Time  3   88.7    29.6   1.813   0.151
```

```
## Residuals    81 1321.0    16.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Varroa Mites through Time by Queen Origin

##	Origin	VarroaDay	n	mean	sd	se
## 1	California	12	20	9.5000000	9.2179492	2.0611961
## 2	California	24	16	0.8666667	0.6399405	0.1599851
## 3	California	47	14	1.4000000	1.2649111	0.3380617
## 4	California	61	9	1.5555556	1.3333333	0.4444444
## 6	Local	12	20	7.4000000	4.8275525	1.0794736
## 7	Local	24	19	0.8947368	0.7374684	0.1691869
## 8	Local	47	17	1.1176471	1.1663165	0.2828733
## 9	Local	61	17	1.5294118	1.6999135	0.4122896



Nosema:

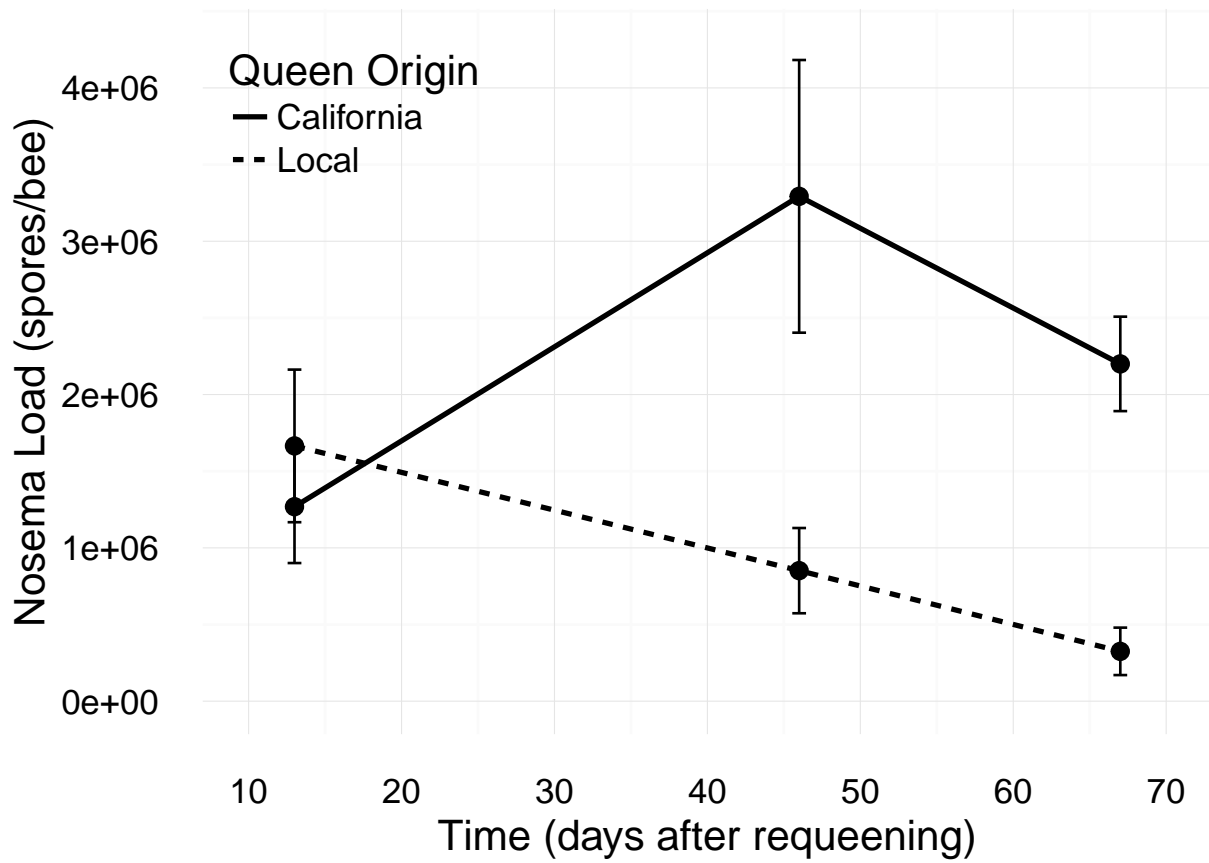
```
# NOSEMA - use repeated measures anova to look at differences between local
# and california through time:
aov.out <- aov(Nosema ~ Origin * Time + Error(FieldID), data=QueenDF)
```

```
# look at summary of model: SIGNIFICANT
summary(aov.out)
```

```
##
## Error: FieldID
##           Df      Sum Sq   Mean Sq F value   Pr(>F)
## Origin      1 3.156e+13 3.156e+13   7.957 0.00794 **
## Time        2 6.900e+12 3.450e+12   0.870 0.42813
## Origin:Time  1 3.155e+11 3.155e+11   0.080 0.77962
## Residuals   34 1.349e+14 3.966e+12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df      Sum Sq   Mean Sq F value   Pr(>F)
## Time        2 9.219e+12 4.609e+12   1.456 0.242438
## Origin:Time  2 5.190e+13 2.595e+13   8.195 0.000792 ***
## Residuals   53 1.678e+14 3.167e+12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Nosema through Time by Queen Origin

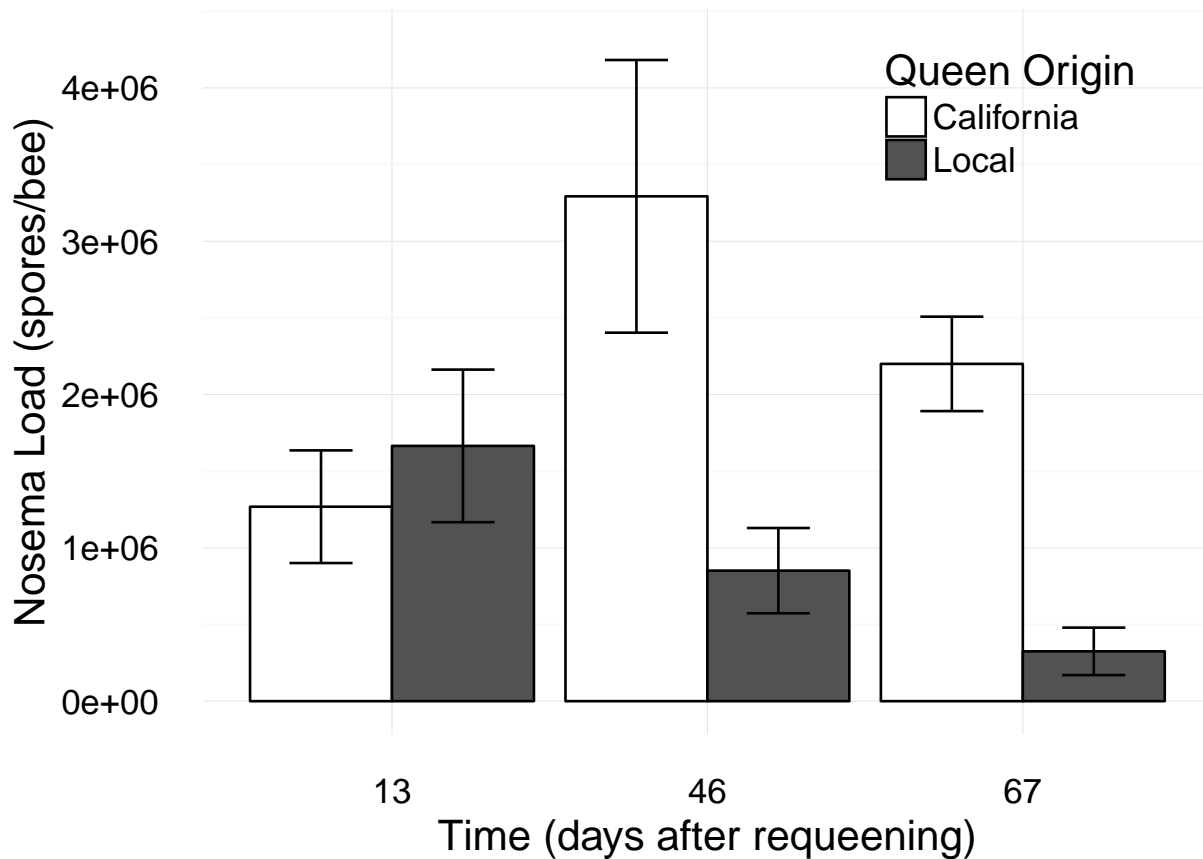
```
##           Origin NosemaDay  n      mean      sd      se
## 1 California      13 20 1268421.1 1643079.8 367403.8
## 2 California      46 14 3292857.1 3327340.5 889269.2
## 3 California      67  9 2200000.0 923901.4 307967.1
## 5      Local      13 20 1665000.0 2225530.7 497643.8
## 6      Local      46 17  851470.6 1146546.1 278078.3
## 7      Local      67 17  325000.0 637806.3 154690.7
```



Nosema through Time by Queen Origin (bar plot)

##	Origin	NosemaDay	n	mean	sd	se
## 1	California	13	20	1268421.1	1643079.8	367403.8
## 2	California	46	14	3292857.1	3327340.5	889269.2
## 3	California	67	9	2200000.0	923901.4	307967.1
## 5	Local	13	20	1665000.0	2225530.7	497643.8
## 6	Local	46	17	851470.6	1146546.1	278078.3
## 7	Local	67	17	325000.0	637806.3	154690.7





## Pollen:

```
# Pollen - use repeated measures anova to look at differences between local
# and california through time:
aov.out <- aov(Pollen ~ Origin * Week + Error(ID), data=PollenDF)

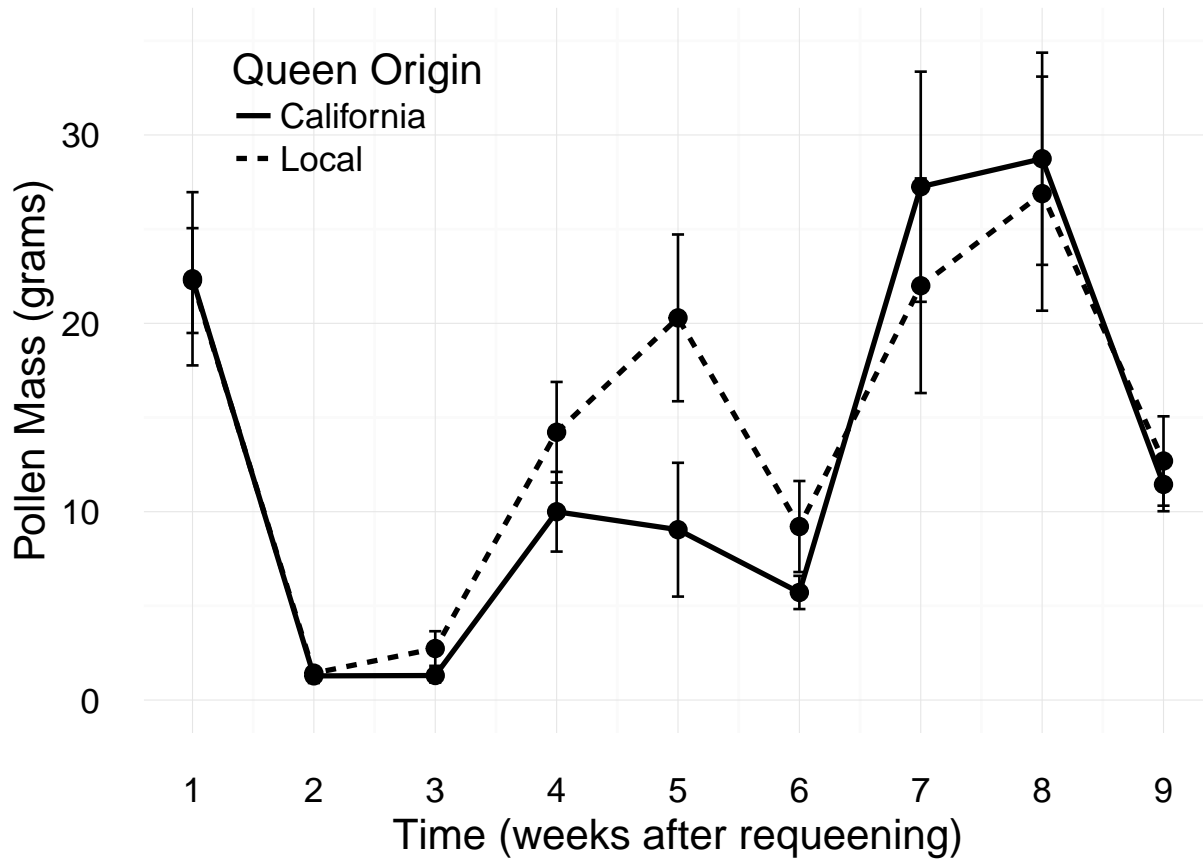
# look at summary of model: NOT SIGNIFICANT
summary(aov.out)
```

```
##
## Error: ID
##          Df Sum Sq Mean Sq F value Pr(>F)
## Origin      1    205    205.2   0.365  0.550
## Week         1    664    663.5   1.180  0.285
## Origin:Week  1    343    343.4   0.611  0.440
## Residuals   35  19677    562.2
##
## Error: Within
##          Df Sum Sq Mean Sq F value Pr(>F)
## Week         1  1206  1206.2   5.084 0.0252 *
## Origin:Week  1    212    211.5   0.892 0.3461
## Residuals   211  50061    237.3
## ---
```

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Weekly Pollen Weight by Queen Origin

##	Origin	Week	n	mean	sd	se
## 1	California	1	20	22.365105	20.566750	4.5988650
## 2	California	2	20	1.284077	1.685422	0.3768718
## 3	California	3	20	1.302333	1.671533	0.3737662
## 4	California	4	20	9.997000	9.460188	2.1153625
## 5	California	5	10	9.044625	11.231272	3.5516402
## 6	California	6	20	5.712222	3.945849	0.8823186
## 7	California	7	20	27.251714	27.324256	6.1098894
## 8	California	8	20	28.737889	25.191477	5.6329854
## 9	California	9	20	11.448444	6.365922	1.4234635
## 10	Local	1	20	22.271053	12.444065	2.7825776
## 11	Local	2	20	1.427316	1.357923	0.3036409
## 12	Local	3	20	2.735313	4.109949	0.9190126
## 13	Local	4	20	14.217444	11.946764	2.6713777
## 14	Local	5	10	20.289900	13.998098	4.4265872
## 15	Local	6	20	9.212882	10.811116	2.4174390
## 16	Local	7	20	21.995933	25.477828	5.6970155
## 17	Local	8	20	26.884750	27.790647	6.2141776
## 18	Local	9	20	12.691235	10.607100	2.3718197



## Viruses

```
# split virus DF by virus for stats:
VirusDF$IntensityMult <- VirusDF$Band_Intensity*106
VirusDFsplit <- split(VirusDF, VirusDF$Virus)

#-----
# DWV - use repeated measures anova to look at differences between local
# and california through time:
aov.out <- aov(IntensityMult ~ Origin * Time + Error(ID), data=VirusDFsplit$DWV)

# look at summary of model: NOT SIGNIFICANT
summary(aov.out)

##
## Error: ID
##           Df      Sum Sq   Mean Sq F value Pr(>F)
## Origin      1 1.352e+12 1.352e+12   0.328  0.571
## Time        1 1.026e+12 1.026e+12   0.249  0.621
## Residuals  36 1.485e+14 4.125e+12
##
## Error: Within
##           Df      Sum Sq   Mean Sq F value Pr(>F)
## Time        1 7.022e+10 7.022e+10   0.016 0.9014
## Origin:Time  1 2.285e+13 2.285e+13   5.084 0.0316 *
## Residuals   30 1.348e+14 4.495e+12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#-----
# BQCV - use repeated measures anova to look at differences between local
# and california through time:
aov.out <- aov(IntensityMult ~ Origin * Time + Error(ID), data=VirusDFsplit$BQCV)

# look at summary of model: NOT SIGNIFICANT
summary(aov.out)

##
## Error: ID
##           Df      Sum Sq   Mean Sq F value   Pr(>F)
## Origin      1 8.614e+11 8.614e+11   2.131    0.153
## Time        1 1.319e+13 1.319e+13  32.631 1.68e-06 ***
## Residuals  36 1.455e+13 4.042e+11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df      Sum Sq   Mean Sq F value   Pr(>F)
## Time        1 8.893e+13 8.893e+13 148.971 2.25e-13 ***
## Origin:Time  1 1.234e+12 1.234e+12   2.067    0.161
## Residuals   31 1.850e+13 5.969e+11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#-----
# IAPV - use repeated measures anova to look at differences between local
# and california through time:
aov.out <- aov(IntensityMult ~ Origin * Time + Error(ID), data=VirusDFsplit$IAPV)

# look at summary of model: SIGNIFICANT
summary(aov.out)
```

```
##
## Error: ID
##           Df      Sum Sq   Mean Sq F value   Pr(>F)
## Origin      1 2.333e+13 2.333e+13  12.153 0.00131 **
## Time        1 1.365e+11 1.365e+11   0.071 0.79129
## Residuals   36 6.911e+13 1.920e+12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df      Sum Sq   Mean Sq F value   Pr(>F)
## Time        1 4.472e+12 4.472e+12   2.828 0.103
## Origin:Time  1 8.007e+11 8.007e+11   0.506 0.482
## Residuals   31 4.902e+13 1.581e+12
```

Viral Load by Virus type at two time points for by queen origin

