Using Big Data to Examine Host-Parasitoid Interactions

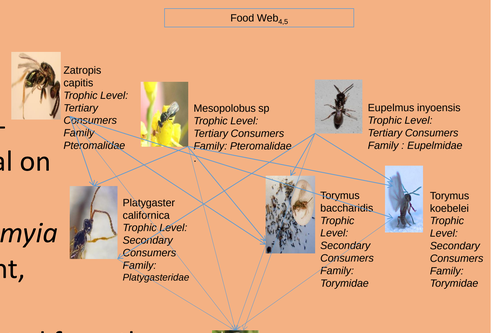
Barnali Dash*, Megan Robertson**

**Mount Holyoke College, **Amherst College*

Introduction

•The experiment and data collection were conducted through the Department of Integrative Biology at the University of California, Berkeley.

•The main motivation for the research was to assess the effect of micro-environment and dispersal on the development and establishment of *Rhopalomyia californica* on its host plant, *Baccharis pilularis.* [1,2]



•The galls that were collected from the experiment were later dissected. However, the dissection data was neither analyzed nor published. The first task therefore was to process the data and make it ready for analysis.

Using R for Data Management

•R is statistical freeware used for analysis and fitting models

•Software used to complete the data management aspects of the project

•Sort data by a variable or view certain galls based on a common characteristic

One of the most important parts of the data management process was the use of metadata.

Plot	Letter
BLP1	A
BLP3	B
BLP4	C
BLP5	D
BLP6	E
CGP1	F
CGP2	G
CGP4	H
CGP5	I

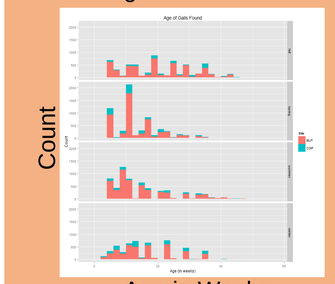
Example of metadata from project

•Metadata
Creation of additional datasets that provide more information about the entire dataset.

•match() function
Adds information from the metadata to the data by matching an identifier between the two sets of data

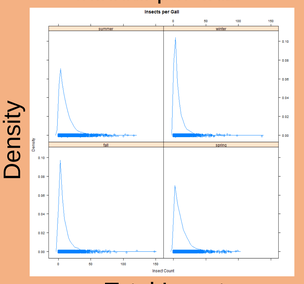
Preliminary Findings

Age of Galls



Age in Weeks

Insects per Gall



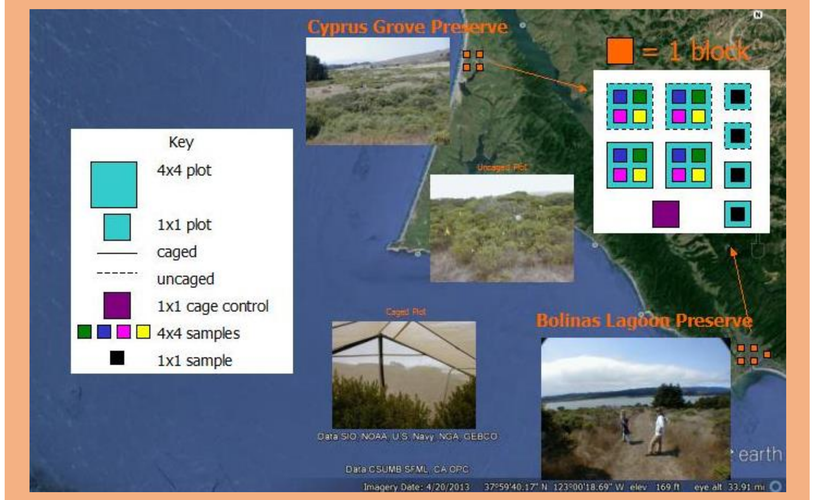
Total Insects

• We are investigating the distribution of gall ages (in weeks) at BLP and CGP.

• From the above figure we notice that while distribution of ages is similar at the two sites, BLP had a lot more galls than CGP.

• The density plot of total insects per gall (per season) confirms previous findings that state that galls formed by *R. californica* tend to contain 1 to >100 insect larvae per gall (with one chamber per larva). [1]

Experimental Design



The experiment took place at two sites:

- Cyprus Grove Preserve — 4 blocks
- Bolinas Lagoon Preserve — 5 blocks

Each block consisted of the following:

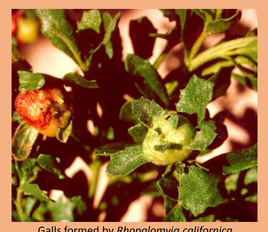
- Four 4x4 plots (2 caged, 2 uncaged) – 4 samples each
- Four 1x1 plots (2 caged, 2 uncaged) – 1 sample each
- One 1x1 cage control plot

Challenges and Difficulties

The data management stage of the project consisted of various tasks necessary to clean the data. There were different challenges faced during this step.

Issues

1. Inconsistent entries for variables (i.e. 'pink' written as PINK, pink, p, pi, etc.)
2. Typos made during data entry
3. Truncation of galls
4. Galls missed during appropriate collection date



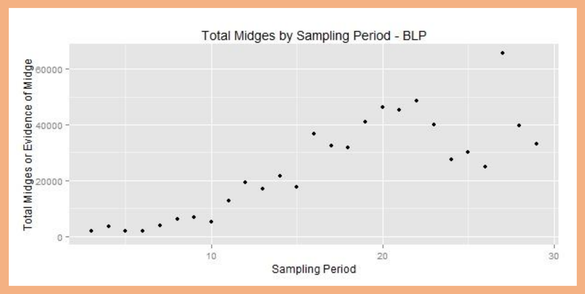
Galls formed by *Rhopalomyia californica*

Platygaster californica

How Issues Were Addressed

1. Metadata and the match function used to make substitutions
2. Handwritten notes utilized to double check entries
3. Incorporate information at analysis stage
4. Complete analysis with and without these galls to see their effect

Future Directions



•Learning more about the relationships and interactions between the populations through time series analysis

•Comparing caged vs. uncaged environments

•Determining the effect of gall truncation

•Simulate theoretical ecological models and compare to populations observed in the data

- Nicholson-Bailey
- Continuous vs. discrete variables
- Dynamic models and state-space models

Sources

1.Boukili, V. K S., M. F. Hoopes, and C. J. Briggs. "Effect of Microenvironment on Development of a Gall Midge." *Environmental Entomology* 36.2 (2007): 441-50. Web. 22 July 2014.

2.Briggs, C.J., and J. Latto. "The Effect of Dispersal on the Population Dynamics of a Gall-forming Midge and Its Parasitoids." *Journal of Animal Ecology* 69.1 (2000): 96-105. Web.

3."California." 37°59'40.17"N" and 123 °00'18.69"W. GOOGLE EARTH. April 20, 2013. July 18, 2014.

4.Digital image. *BugGuide.Net*. N.p., n.d. Web. 18 July 2014. <http://bugguide.net/>.

5.Force, D. C. "Ecology of Insect Host-Parasitoid Communities." *Science* 184.4137 (1974): 624-32. Web.

6.R Development Core Team: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2009. ISBN 3-900051-07-0