Aidan Coyle

QERM 514

Final report

Modeling *Hematodinium sp.* infection in Alaskan Tanner crab (*Chionoecetes bairdi*)

**Introduction:**

*Hematodinium sp.* (hereafter *Hematodinium*) is a globally-distributed parasitic dinoflagellate that infects a wide range of crustaceans. Some of its most commercially-important hosts are Alaskan snow and Tanner crab (*Chionoecetes opilio* and *Chionoecetes bairdi*). Within *Chionoecetes* crab, infections develop slowly, over a period of several months. As *Hematodinium* multiplies in the hemolymph, the crab takes on a pale white color and its flesh becomes chalky and bitter. Hence, *Hematodinium* infection is colloquially known as bitter crab syndrome (BCS). Over time, the host becomes more lethargic as the parasite multiplies further, eventually resulting in death.

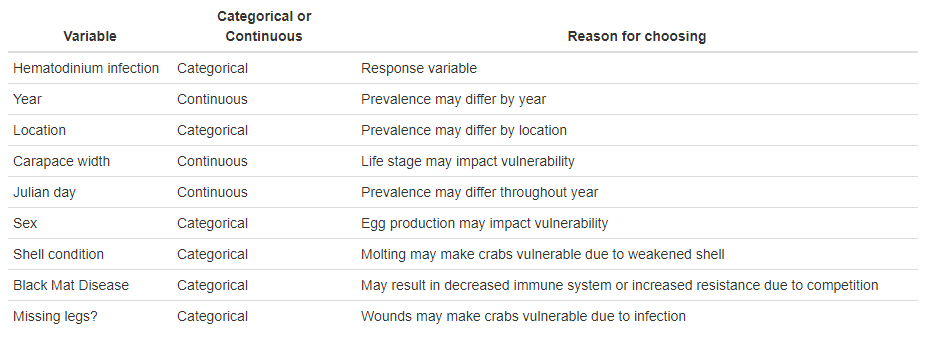
Determining the dynamics of *Hematodinium* infection and the factors associated with its spread is vitally important for proper management of Alaskan *Chionoecetes* fisheries. Unfortunately, there is still a great deal that is unknown. Currently, most researchers operate from the assumption that there is an infective waterborne stage that penetrates through the shell, and thus that crab are most vulnerable when the shell is compromised in some form, either through molting or through some injury to the crab. However, this is far from proven. The relationship of *Hematodinium* with other biological and ecological factors with *Hematodinium* infection is currently undetermined.

**Methods**

The Alaska Department of Fish and Game has been conducting annual pot surveys of the southeast Alaska Tanner crab population for several decades. These surveys, which are typically conducted between June and October, consist of six core stations and seven peripheral stations. Our data set includes six years of survey data, beginning in 2007 and ending in 2012. Surveys occurring on even-numbered years included the core and peripheral stations, while surveys from odd-numbered years only sampled the core stations. Each row contains data on a single crab, including information on the time and place of its catch and biological information on the crab. Importantly, *Hematodinium* infection is visually assessed rather than tested. Due to the slow progression of infection, crab only become visually positive long after their initial infection – potentially months after.

We planned to illuminate the factors that may impact the prevalence and distribution of Tanner crab infected with *Hematodinium* with the use of modeling. We pulled columns from the survey data that could be explanatory to either *Hematodinium* infection status or distribution (Table 1).

Prior to model creation, we cleaned and checked our data. All lines containing NAs were removed, along with all lines describing other crab species. Apparent data entry errors were also removed. We then checked for correlation between our variables. Pearson’s test was used to check for correlation between continuous variables, while Cramer’s V test was used to check for



**Table 1:** Variables included in model analysis and reasoning.

correlation between categorical variables. Finally, Spearman rank-order correlation was used to determine whether categorical and continuous variables were correlated. Unsurprisingly, as mature males are much larger than females, we found a correlation between carapace width and sex. Therefore, when creating our models, we determined that no model would include both variables. No other correlations > 0.6 were observed.

Through previous discussion with scientists working in this system, we gathered that *Hematodinium* prevalence is consistently different among locations, and we assume that its prevalence varies by year. Furthermore, our data are Bernoulli-distributed, with each crab either infected or uninfected. Therefore, generalized linear mixed models (GLMMs) were used to model our data, with random effects for year and location. However, a full model was too large to run directly using Laplace approximation. To reduce complexity, we first scaled all continuous variables and reduced the categories of our categorical variables when possible. Then, using the R package MASS, we built two full models – one including all but sex, the other all but carapace width - using penalized quasi-likelihood (PQL). Both had nearly-identical parameter estimates. Fixed effects were ranked according to the absolute value of their coefficients, and the largest two (Black Mat disease and shell condition) were added to a new GLMM built using Laplace approximation along with the random effects. Fixed effects were then added to the model using the update() function. Through this, we were able to create two near-full models – again, one excluding sex and the other excluding carapace width. Both excluded missing leg status, as it had the smallest coefficient in our PQL models. To verify it was acceptable to drop missing leg status, we built a third model that included all variables except carapace width (its opposite – all but sex – would not run). The AICc of the model that included missing leg status was higher than the model excluding it, confirming our decision. To decide which of our correlated variables to include, we then compared the AICc of our two near-full models, and found that the model that included carapace width was superior. Therefore, we decided to move forward with a full model that included all variables but sex and missing leg status.

We then used the dredge() function from the R MuMIn package to test all combinations of fixed effects in our full model, compare using AICc, and weight accordingly. This produced two models with weights above zero, which we then averaged using MuMIn::model.avg(). This average model was concluded to be our optimal model. We then tested our model using model diagnostics TK TK TK TK

**Results:**

The average model included five variables (Table 2). Of these, three were significant by p-value – depth, shell condition, and width. Of these, shell condition had the largest coefficient, followed by depth and then width. Interestingly, despite not being statistically significant, Black Mat disease status had by far the largest effect on the model

