

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, including Alaskan snow and Tanner crab (*Chionoecetes opilio* and *Chionoecetes bairdi*). For *Chionoecetes* crabs, infections develop 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 (and thus unmarketable). Over time, lethargy increases as the parasite multiplies, 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, a great deal is still unknown. Currently, most researchers operate from the assumption that there is an infective waterborne stage that penetrates through the shell, and thus 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 not yet proven. The relationship of *Hematodinium* infection with other biological and ecological factors is currently undetermined.

Methods

The Alaska Department of Fish and Game (ADF&G) has conducted annual pot surveys of the southeast Alaska Tanner crab population for several decades. These surveys, which occur

between June and October, consist of six core stations and seven peripheral stations. Our dataset includes six years of survey data, beginning in 2007 and ending in 2012. Surveys occurring on even-numbered years included core and peripheral stations, while surveys from odd-numbered years only sampled core stations. Each row contains data on a single crab, including the time and place of its catch and biological information. Importantly, *Hematodinium* infection is visually assessed rather than tested. Due to the slow progression of infection, crab only become visually positive potentially months after their infection.

We used modeling to illuminate the factors that may impact the prevalence and distribution of Tanner crab infected with *Hematodinium*. We pulled columns from the survey data that could explain either *Hematodinium* infection status or distribution (Table 1). Prior to model creation, we cleaned and checked our data. Lines containing NAs were removed, along

Variable	Categorical or Continuous	Reason for choosing	AIC compared to null model
Hematodinium infection	Categorical	Response variable	NA
Year	Continuous	Random effect. Prevalence may differ by year	NA
Location	Categorical	Random effect. Prevalence may differ by location	NA
Shell condition	Categorical	Molting may make crabs vulnerable due to weakened shell	-791.8
Carapace width	Continuous	Life stage may impact vulnerability	-104.2
Black Mat Disease	Categorical	May weaken immune system or compete with <i>Hematodinium</i>	-73.8
Sex	Categorical	Egg production may impact vulnerability	-19.4
Missing legs?	Categorical	Wounds may make crabs vulnerable due to infection	-8.9
Julian day	Continuous	Prevalence may differ throughout year	-3.3

Table 1: Variables included in model analysis and reasoning. Order corresponds to importance in full model. A model for each fixed effect was created, containing only the response variable, the random effects, and the fixed effect. Each model was then compared, using AIC, to a null model containing only the response variable and random effects.

with 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 categorical variables, and Spearman rank-order correlation was used to test categorical and continuous variables together. Unsurprisingly, we found a correlation between carapace width and sex, as mature males are much larger than females. Therefore, we determined no full model would include both variables. No other correlations > 0.6 were observed.

Through previous discussion, we gathered that *Hematodinium* prevalence consistently differs among locations, and we assume that its prevalence varies between years. Furthermore, our data are Bernoulli-distributed, as each crab is 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 scaled continuous variables and reduced the categories of categorical variables when appropriate. We then built a series of models using Laplace approximation. Each included one fixed effect plus our random effects. We then compared the AIC of each model to a null model (Table 1). This estimated the importance of each fixed effect. We then slowly assembled a larger model by adding fixed effects in order of importance. This created two near-full models, one excluding sex and one excluding carapace width. Both excluded missing leg status, as it was the second-least important and broke the model when added. The least important variable (day) was added to the "full" models, as it still reduced AIC from the null model without impeding model creation. To decide which correlated variable to include, we compared the AICc of our two near-full models, and found the model that included carapace width was superior. Therefore, we moved 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.

Results:

After building our optimal model, we tested goodness of fit using Pearson's chi-square test and plotted Q-Q plots of residuals. The model passed both diagnostics. However, it failed the Hosmer-Lemeshow test, indicating that the data include variation that isn't accounted for in the model. In other words, the model is missing at least one important parameter. Due to this, results from this model should be treated with a great deal of caution, and should not be seen as necessarily biologically meaningful. That said, they still contain interesting information. The average model included five variables (Table 2). Of these, three were significant – depth, shell condition, and width. The presence of Black Mat disease and the day of catch were not significant.

	Estimate	Std. Error	Adjusted SE	z value	Pr (> z)
(Intercept)	-10.65688	11.85864	11.85963	0.899	0.369
BLACKMAT_CODE.L	-10.12176	16.77164	16.77304	0.603	0.546
DEPTH_SCALED	-0.50887	0.05463	0.05464	9.313	<2e-16 ***
SHELL_CONDITION_CODE.L	-1.54969	0.26261	0.26263	5.901	<2e-16 ***
SHELL_CONDITION_CODE.Q	-1.68320	0.19912	0.19914	8.453	<2e-16 ***
SHELL_CONDITION_CODE.C	0.64983	0.10591	0.10592	6.135	<2e-16 ***
WIDTH_SCALED	0.20714	0.03107	0.03108	6.665	<2e-16 ***
DAY_SCALED	-0.00975	0.05200	0.05200	0.187	0.851

Table 2: Summary of the optimal model constructed by averaging the two models of *Hematodinium* infection with weights above 0

Discussion:

Though our model is incomplete and our missing parameter(s) could covary with any of our existing variables, we do have some intriguing results from our average model. Notably, shell condition is highly significant (Table 2). If we use the ADF&G conversion charts between shell condition and time since molt, crab appear to become infected during molt, infections develop over several months into visual positives, followed by eventual mortality (Fig. 1). Interestingly, shell vulnerabilities also occur when legs are lost, but missing leg status was excluded from our model. This is yet another piece of evidence linking infection to molting, and suggests that lab experiments, potentially involving the exposure of freshly molted crab to *Hematodinium*, could be a useful way to determine whether this when infection occurs.

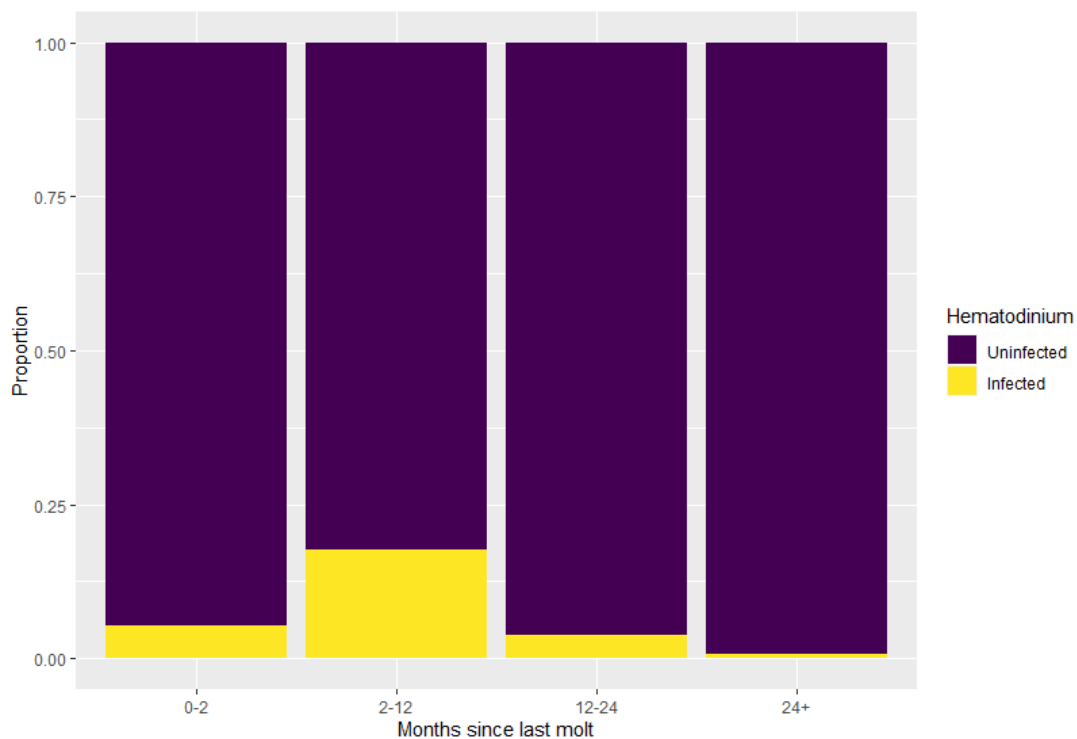


Figure 1: Hematodinium infection rates and time since last molt. Shell condition was converted to months since last molt using ADF&G conversion table from survey protocol. Visible infection rates are highest from 2-12 months post-molt, followed by a decrease over time, presumably due to crab mortality.

Despite their high significance, the effects of depth and carapace width are more muted (Table 2). However, larger crabs appear to have a slightly higher rate of infection. This indicates that older crabs could be more vulnerable. *Hematodinium* infections are also slightly more prevalent at shallow depths. Due to the long incubation period, this is likely due to the depth preferences of infected crab rather than infection being more prevalent at shallower depths. Perhaps weakened crab prefer shallower waters due to other factors, such as food availability, or perhaps *Hematodinium* drives the crab towards shallower waters – as we see in rhizocephalan barnacles, crab parasites are certainly capable of altering behavior. This preference indicates an optimal fishing strategy could take depth into account, as crab with developed infections are not kept.

Though not significant in the model, Black Mat syndrome was included, and yields interesting results. The standard error was extremely high (Table 2), potentially due to the small sample size – Black Mat disease has a low baseline prevalence and disproportionately affects older shell-condition crab, so even in our large dataset, we would expect few crabs with both Black Mat and *Hematodinium*. However, despite sampling over 700 crabs with Black Mat, none were found with both diseases. Therefore, larger data sets could prove useful in determining whether the fungus causing Black Mat disease prevents *Hematodinium* infections from occurring or developing.

The Hosmer-Lemeshow test revealed our model is missing at least one important predictor. Unfortunately, as our model included all possible relevant parameters, this means our dataset is inadequate. Two parameter varieties seem to be most likely – predator abundance and environmental factors. Of the environmental variables, temperature and salinity are particularly strong candidates, as they have been linked to *Hematodinium* prevalence in other systems.

125 Unfortunately, the ADF&G Tanner crab survey data has information on neither. However, the
126 National Oceanic and Atmospheric Administration (NOAA) performs an annual bottom trawl
127 survey of the Eastern Bering Sea (EBS). That dataset contains all variables from this model,
128 along with information on temperature and the prevalence of potential predators. Furthermore,
129 salinity is much more homogenous within the Bering Sea than within the inlets of southeastern
130 Alaska. Therefore, creating a model for the EBS dataset could yield more substantive insights
131 into *Hematodinium* dynamics within Alaskan Tanner crab, and perhaps uncover which
132 parameters are missing from our model.