1   Aidan Coyle

2   QERM 514

3   Final report

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

5   **Introduction:**

6   *Hematodinium sp.* (hereafter *Hematodinium*) is a globally distributed parasitic

7   dinoflagellate that infects a wide range of crustaceans, including Alaskan snow and Tanner crab

8   (*Chionoecetes opilio* and *Chionoecetes bairdi*). For *Chionoecetes* crabs, infections develop over

9   a period of several months. As *Hematodinium* multiplies in the hemolymph, the crab takes on a

10 pale white color and its flesh becomes chalky and bitter (and thus unmarketable). Over time,

11 lethargy increases as the parasite multiplies, eventually resulting in death.

12   Determining the dynamics of *Hematodinium* infection and the factors associated with its

13 spread is vitally important for proper management of Alaskan *Chionoecetes* fisheries.

14 Unfortunately, a great deal is still unknown. Currently, most researchers operate from the

15 assumption that there is an infective waterborne stage that penetrates through the shell, and thus

16 crab are most vulnerable when the shell is compromised in some form, either through molting or

17 through some injury to the crab. However, this is not yet proven. The relationship of

18 *Hematodinium* infection with other biological and ecological factors is currently undetermined.

19   **Methods**

20   The Alaska Department of Fish and Game (ADF&G) has conducted annual pot surveys

21 of the southeast Alaska Tanner crab population for several decades. These surveys, which occur

22   between June and October, consist of six core stations and seven peripheral stations. Our dataset

23   includes six years of survey data, beginning in 2007 and ending in 2012. Surveys occurring on

24   even-numbered years included core and peripheral stations, while surveys from odd-numbered

25   years only sampled core stations. Each row contains data on a single crab, including the time and

26   place of its catch and biological information. Importantly, *Hematodinium* infection is visually

27   assessed rather than tested. Due to the slow progression of infection, crab only become visually

28   positive potentially months after their infection.

29   We used modeling to illuminate the factors that may impact the prevalence and

30   distribution of Tanner crab infected with *Hematodinium*. We pulled columns from the survey

31   data that could explain either *Hematodinium* infection status or distribution (Table 1). Prior to

32   model creation, we cleaned and checked our data. Lines containing NAs were removed, along

33

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 Hematodinium	-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.

41   with lines describing other crab species. Apparent data entry errors were also removed. We then

42   checked for correlation between our variables. Pearson's test was used to check for correlation

43 between continuous variables, while Cramer's V test was used to check for categorical variables,

44 and Spearman rank-order correlation was used test categorical and continuous variables together.

45 Unsurprisingly, we found a correlation between carapace width and sex, as mature males are

46 much larger than females. Therefore, we determined no full model would include both variables.

47 No other correlations > 0.6 were observed.

48   Through previous discussion, we gathered that *Hematodinium* prevalence consistently

49 differs among locations, and we assume that its prevalence varies between years. Furthermore,

50 our data are Bernoulli-distributed, as each crab is either infected or uninfected. Therefore,

51 generalized linear mixed models (GLMMs) were used to model our data, with random effects for

52 year and location. However, a full model was too large to run directly using Laplace

53 approximation. To reduce complexity, we scaled continuous variables and reduced the categories

54 of categorical variables when appropriate. We then built a series of models using Laplace

55 approximation. Each included one fixed effect plus our random effects. We then compared the

56 AIC of each model to a null model (Table 1). This estimated the importance of each fixed effect.

57 We then slowly assembled a larger model by adding fixed effects in order of importance. This

58 created two near-full models, one excluding sex and one excluding carapace width. Both

59 excluded missing leg status, as it was the second-least important and broke the model when

60 added. The least important variable (day) was added to the "full" models, as it still reduced AIC

61 from the null model without impeding model creation. To decide which correlated variable to

62 include, we compared the AICc of our two near-full models, and found the model that included

63 carapace width was superior. Therefore, we moved forward with a full model that included all

64 variables but sex and missing leg status.

65   We then used the dredge() function from the R MuMIn package to test all combinations

66   of fixed effects in our full model, compare using AICc, and weight accordingly. This produced

67   two models with weights above zero, which we then averaged using MuMIn::model.avg(). This

68   average model was concluded to be our optimal model.

69   **Results:**

70   After building our optimal model, we tested goodness of fit using Pearson's chi-square

71   test and plotted Q-Q plots of residuals. The model passed both diagnostics. However, it failed the

72   Hosmer-Lemeshow test, indicating that the data include variation that isn't accounted for in the

73   model. In other words, the model is missing at least one important parameter. Due to this, results

74   from this model should be treated with a great deal of caution, and should not be seen as

75   necessarily biologically meaningful. That said, they still contain interesting information. The

76   average model included five variables (Table 2). Of these, three were significant – depth, shell

77   condition, and width. The presence of Black Mat disease and the day of catch were not

78   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

79

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

81   **Discussion:**

82   Though our model is incomplete and our missing parameter(s) could covary with any of

83   our existing variables, we do have some intriguing results from our average model. Notably,

84   shell condition is highly significant (Table 2). If we use the ADF&G conversion charts between

85   shell condition and time since molt, crab appear to become infected during molt, infections

86   develop over several months into visual positives, followed by eventual mortality (Fig. 1).

87   Interestingly, shell vulnerabilities also occur when legs are lost, but missing leg status was

88   excluded from our model. This is yet another piece of evidence linking infection to molting, and

89   suggests that lab experiments, potentially involving the exposure of freshly molted crab to

90   *Hematodinium*, could be a useful way to determine whether this when infection occurs.

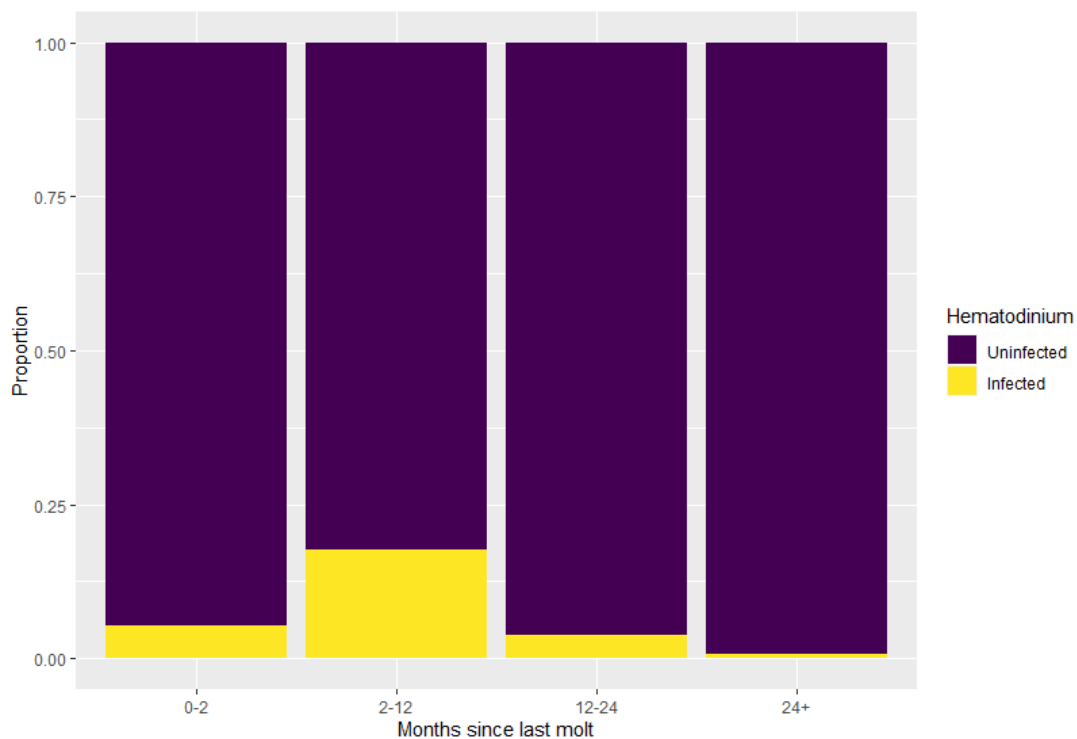91

92

93

94

95

96

97

98

99

100

Figure 1: Hematodinium infection rates and time since last molt (as determined by shell condition). Visible infection rates are highest from 2-12 months post-molt, followed by a decrease over time, presumably due to crab mortality.

101

102   Despite their high significance, the effects of depth and carapace width are more muted

103   (Table 2). However, larger crabs appear to have a slightly higher rate of infection. This indicates

104   that older crabs could be more vulnerable. *Hematodinium* infections are also slightly more

105   prevalent at shallow depths. Due to the long incubation period, this is likely due to the depth

106   preferences of infected crab rather than infection being more prevalent at shallower depths.

107   Perhaps weakened crab prefer shallower waters due to other factors, such as food availability, or

108   perhaps *Hematodinium* drives the crab towards shallower waters – as we see in rhizocephalan

109   barnacles, crab parasites are certainly capable of altering behavior. This preference indicates an

110   optimal fishing strategy could take depth into account, as crab with developed infections are not

111   kept.

112   Though not significant in the model, Black Mat syndrome was included, and yields

113   interesting results. The standard error was extremely high (Table 2), potentially due to the small

114   sample size – Black Mat disease has a low baseline prevalence and disproportionately affects

115   older shell-condition crab, so even in our large dataset, we would expect few crabs with both

116   Black Mat and *Hematodinium*. However, despite sampling over 700 crabs with Black Mat, none

117   were found with both diseases. Therefore, larger data sets could prove useful in determining

118   whether the fungus causing Black Mat disease prevents *Hematodinium* infections from occurring

119   or developing.

120   The Hosmer-Lemeshow test revealed our model is missing at least one important

121   predictor. Unfortunately, as our model included all possible relevant parameters, this means our

122   dataset is inadequate. Two parameter varieties seem to be most likely – predator abundance and

123   environmental factors. Of the environmental variables, temperature and salinity are particularly

124   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.