

# Introduction of CRISPR-cas9 gene into invasive Lion fish population

Tom Eichlersmith

This mathematical model is designed to simulate how the lion fish population in the Caribbean will react to the introduction of a gene that renders female offspring infertile. With the introduction of this new genotype, there will be three different characteristics of each individual that this model focuses on: stage, sex, and genotype. The different options are

- Stage: Larval (L), Juvenile (J), Adult (A)
- Sex: Male (m), Female (f)
- Genotype: Wild (w), CRISPR Carrier (c), Infertile (i)

We note here that only female lion fish are able to be rendered infertile by this gene. The only stage that is able to reproduce is Adults and the effects of the genotypes of the parents are given by

- $A_{cf}$  mating with  $A_{wm}$  or  $A_{cm}$  produces  $L_{if}$  and  $L_{cm}$
- $A_{cm}$  mating with  $A_{wf}$  produces  $L_{cf}$  and  $L_{cm}$
- $A_{wm}$  mating with  $A_{wf}$  produces  $L_{wf}$  and  $L_{wm}$

There are multiple ways to represent these pairings, but I have chosen to focus on the populations of  $A_{cf}$ ,  $A_{cm}$ , and  $A_{wm}$  because it simplifies the matrix (as you will see below).

This model is an expansion of the one given by Morris et. al. in [1]. There, the model assumes only Wild genotypes.

---

## Matrix Model

In order to keep the same form of the matrix model from [1] and introduce various different genotypes for each of the stages, I expanded the matrix to have sub-matrices and the population vector to have sub-vectors. This will mean that the matrix itself will end up have a much larger size and will contain many zeros, but it will look pretty so I'm doing it.

The model is iterated by the equation

$$\begin{pmatrix} \mathbf{L}(t+1) \\ \mathbf{J}(t+1) \\ \mathbf{A}(t+1) \end{pmatrix} = \begin{pmatrix} 0 & 0 & f e^{-M_E D_E} \mathbf{SP} \\ G_L \mathbf{1} & \frac{11}{12} e^{-M_J} \mathbf{1} & 0 \\ 0 & \frac{1}{12} e^{-M_J} \mathbf{1} & P_A \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{H} & 0 \\ 0 & 0 & \mathbf{H} \end{pmatrix} \left( \begin{pmatrix} \mathbf{L}(t) \\ \mathbf{J}(t) \\ \mathbf{A}(t) \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{C}_J \\ 0 \end{pmatrix} \right)$$

where some elements of the matrices explicitly depend upon the time step as well as the population vectors of the juvenile and adult stages, and expressions for the elements are given below.  $\mathbf{1}$  is the 5x5 identity matrix and 0 is used interchangeable as the 5x5 zero matrix and the 5-dimensional zero vector.

$$\mathbf{L} = \begin{pmatrix} L_{cf} \\ L_{cm} \\ L_{wf} \\ L_{wm} \\ L_{if} \end{pmatrix}, \mathbf{J} = \begin{pmatrix} J_{cf} \\ J_{cm} \\ J_{wf} \\ J_{wm} \\ J_{if} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} A_{cf} \\ A_{cm} \\ A_{wf} \\ A_{wm} \\ A_{if} \end{pmatrix}$$

where  $Q_{TS}$  is the population of lion fish that are type T, sex S, and in stage Q. If we let the proportion of wild type females be defined as  $r_{wf} = A_{wf} / (A_{cf} + A_{wf} + A_{if})$ , then

$$\mathbf{P} = \begin{pmatrix} 0 & r_{wf} R_f & 0 & 0 & 0 \\ 1 - R_f & r_{wf}(1 - R_f) & 0 & 0 & 0 \\ 0 & 0 & 0 & r_{wf} R_f & 0 \\ 0 & 0 & 0 & r_{wf}(1 - R_f) & 0 \\ R_f & 0 & 0 & 0 & 0 \end{pmatrix}$$

represents the production of the different lion fish pairings. If we let  $\eta$  be the probability that the CRISPR gene is not passed on (when it should have been), then

$$\mathbf{S} = \begin{pmatrix} 1 - \eta & 0 & 0 & 0 & 0 \\ 0 & 1 - \eta & 0 & 0 & 0 \\ \eta & 0 & 1 & 0 & \eta \\ 0 & \eta & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 - \eta \end{pmatrix}$$

represents the “switching” of genotypes immediately following birth in order to replicate when the gene is not passed on. If we let the total population be defined as

$T = J_{cf} + J_{cm} + J_{wf} + J_{wm} + J_{if} + A_{cf} + A_{cm} + A_{wf} + A_{wm} + A_{if}$ , then

$$\mathbf{H} = \mathbf{1} - \frac{h(t, T)}{T} \mathbf{1} = \begin{pmatrix} 1 - \frac{h(t, T)}{T} & 0 & 0 & 0 & 0 \\ 0 & 1 - \frac{h(t, T)}{T} & 0 & 0 & 0 \\ 0 & 0 & 1 - \frac{h(t, T)}{T} & 0 & 0 \\ 0 & 0 & 0 & 1 - \frac{h(t, T)}{T} & 0 \\ 0 & 0 & 0 & 0 & 1 - \frac{h(t, T)}{T} \end{pmatrix}$$

represents the harvesting of lion fish. Also we define

$$\mathbf{C}_J = \begin{pmatrix} R_f c(t) \\ (1 - R_f) c(t) \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

to represent the introduction of juvenile lion fish that are carriers of the infertility gene. With the definitions of the matrices and vectors out of the way, I turn now to the long list of elements of these matrices and what those elements represent. I will start with the constants that are unchanged from [1].

- $f = 194577$  (the fecundity)
- $M_E = 0.31$  (egg mortality in days<sup>-1</sup>)
- $D_E = 3$  (egg duration in days)
- $G_L = e^{-M_L D_L}$

- $M_L = 0.35$  (larval mortality in days<sup>-1</sup>)
- $D_L = 30$  (larval duration in days)
- $P_A = e^{-M_A}$
- $M_A = 0.052$  (adult mortality in months<sup>-1</sup>)
- $M_{J0} = 0.165$  (juvenile mortality without density dependence in months<sup>-1</sup>)
- $R_f = 0.46$  (proportion of the population that is female)

Now I turn to the new variables that this model introduces

- $M_J = M_{J0} + \mu T$  (juvenile mortality with density dependence in months<sup>-1</sup>)
  - $T$  is the total population of juveniles and adults defined above
  - $\mu$  is the “mortality slope”, a linear relation between population and juvenile mortality is assumed
- $h(t, T)$  is the function determining the number of lion fish harvested during time step  $t$ . Dependence on total population  $T$  is included. For this model the function is given by

$$h(t, T) = \begin{cases} 0 & \text{if } t < b_h \text{ or } t > b_h + l_h \\ \min\{h_0, \psi T\} & \text{otherwise} \end{cases}$$

- $b_h$  time step to begin harvesting,  $l_h$  number of time steps to harvest,  $h_0$  number of lion fish to attempt to harvest,  $\psi$  fraction of total population that is “catch-able”
- $c(t)$  is the function determining the number of juveniles carrying the infertility gene to add to the population during time step  $t$ . For this model the function is given by

$$c(t) = \begin{cases} 0 & \text{if } t < b_c \text{ or } t > b_c + l_c \\ c_0 & \text{otherwise} \end{cases}$$

- $b_c$  time step to begin releasing,  $l_c$  number to time steps to release,  $c_0$  number of lion fish to release each time step
- $\eta$  is the probability that the CRISPR gene is *not* successfully passed on to a certain offspring from a parent (if  $\eta = 0$ , then the model is assuming that the CRISPR gene is perfectly passed on to all offspring of CRISPR parents)

These are a lot of different values to keep track of, so I am going to separate them into lists depending on who (Nature vs Humans) directly controls their value

- Nature Controlled:  $f, M_E, D_E, M_L, D_L, M_A, R_f, M_{J0}, \mu, \psi, \eta$
- Human Controlled:  $b_h, l_h, h_0, b_c, l_c, c_0$

The values that this model focuses on are the human controlled values (for obvious reasons) as well as the nature controlled mortality slope  $\mu$ , the probability of not passing on the gene  $\eta$ , and fraction of population that is “catch-able”  $\psi$  (because these nature controlled values do not have other literature backing and therefore could take on a wider range of values). The model developed in R is meant to have the flexibility necessary to evaluate how the lion fish population depends upon these eight different variables.

---

## References

- [1]: Morris, Shertzer, and Rice. *A stage-based matrix population model of invasive lionfish with implications for control*. November 2009. DOI 10.1007/s10530-010-9786-8.