Daisy Quorum Gene Drive

Jackie Chea

Math 243, Spring 2018

Abstract

CRISPR-based gene drive could potentially solve many problems by propagating engineered alleles through a wild population, but even the least effective drive is highly invasive to non-target populations. “Daisy quorum gene drive” has been proposed as a potential system for localized spread. We have found a quorum system produces a threshold drive, meaning that the number of engineered individuals released must be above the threshold to achieve spread, and introducing a daisy element decreases the threshold, making the system more practical. Most importantly, we demonstrated that when the migration rate between two populations is low, the engineered alleles can be contained in the population in which they were released.

Introduction

There are many problems in public health, agriculture, and ecology that could be solved by engineering wild populations to remove undesirable phenotypes. For example, by making wild mosquitoes unable to transmit diseases, we could potentially eradicate malaria from the world.[1, 2] We could also eliminate pesticide and herbicide resistance in insects and weeds to improve agricultural yield or render invasive species sterile.[2]

Unfortunately, modifying alleles tends to lower their evolutionary fitness, resulting in the engineered alleles being eliminated by natural selection. One exciting potential solution is gene drive based on RNA-guided CRISPR-Cas9 nucleases. At a locus with one engineered allele and one wild allele, a nuclease can be directed to cut the wild allele. Through homologous repair, the engineered allele will be copied in its place, changing a heterozygote to an engineered homozygote.[1, 2] This creates non-Mendelian inheritance that is biased in favor of the engineered alleles.

Gene drive creates rapid spread of engineered alleles throughout a target population. Consequently, it is difficult to contain the spread of the engineered alleles once they are released. Even considering the likely development of drive-resistant alleles, even the least effective drive will be highly invasive to non-target populations connected through gene flow.[3]

“Daisy quorum gene drive” has been presented as a potential system for ensuring that the wild population would only be altered within a localized area.[4] It consists of two aspects: quorum and daisy.

For the quorum aspect, let there be 2 haploinsufficient essential genes, A and B. We can create an allele A* that is equivalent in function to A, but is inserted at the B locus, and an allele B* that is equivalent in function to B, but is inserted at the A locus. In the dynamics of a quorum

system, there are only two stable equilibria: 100% wild-type homozygote, and 100% wild-type engineered homozygote. Consequently, there will be a threshold such that a certain number of engineered individuals must be released to achieve spread of the engineered alleles. Otherwise, the engineered alleles will go extinct. This threshold behavior facilitates containment.

Daisy drive is a form of drive in which genetic elements are arranged in a daisy-chain so that each element drives the next.[5] This temporarily boosts the frequency of the last engineered allele in the chain. Once the first drive allele is diluted by normal inheritance patterns, the frequencies of the engineered alleles decline.

To combine both aspects into daisy quorum gene drive, we have the engineered C drive allele that cuts A and B alleles to increase the frequency of the engineered B* and A* alleles carrying the desired cargo elements such as disease resistance. The quorum aspect should create a threshold effect for localized spread, and the daisy aspect should lower the threshold, increasing the practicality of the system.

Model

Daisy Quorum Gene Drive in One Population

To run simulations in Matlab, we created a replicator-based model as follows.

In our system, there are three loci, so each gamete's haplotype can be described by $a = (a_1, a_2, a_3)$. a_1 is the daisy locus C, a_2 is the cargo-carrying quorum locus B, and a_3 is the cargo-carrying quorum locus A. At each locus is either the wild-type allele, $a_i = 0$, or the engineered allele, $a_i = 1$. For example, $WBB^* = (0, 0, 1)$. There are 8 total haplotypes. Each genotype can be represented as a matrix of two gamete vectors, $[a, b] = [(a_1, a_2, a_3), (b_1, b_2, b_3)]$. At each of the three loci are one of three potential genotypes: $a_i, b_i = 0, 0$, $a_i, b_i = 0, 1$, or $a_i, b_i = 1, 1$, so there are 27 total genotypes.

Collecting individuals into a population, x_{ab} is the frequency of individuals with genotype $[a, b]$, and f_{ab} is their fitness. The relative rate of production of gametes with haplotype a is represented by:

$$g_a = \sum_{\alpha\beta} x_{\alpha\beta} f_{\alpha\beta} P_{\alpha\beta,a}.$$

The probability an individual with genotype αβ produces a gamete with haplotype *a* is represented by:

$$P_{\alpha\beta,a} = \frac{\delta_{\alpha_1 a_1} + \delta_{\beta_1 a_1}}{2} \prod_{i=2}^{3} \left\{ \delta_{\alpha_1 0}\delta_{\beta_1 0} \frac{\delta_{\alpha_i a_i} + \delta_{\beta_i a_i}}{2} + \left(1 - \delta_{\alpha_1 0}\delta_{\beta_1 0}\right) \times \left[\delta_{\alpha_i \beta_i}\delta_{\alpha_i a_i} + \left(1 - \delta_{\alpha_i \beta_i}\right) \frac{1 + H(-1)^{a_i+1}}{2} \right] \right\}$$

.

δ_{ab} is the Kronecker delta, defined as 1 if *a* = *b* and 0 otherwise. There is one of these equations for each genotype/haplotype pair (27 genotypes * 8 haplotypes = 216 equations). The parameter *H* is the *homing efficiency*, or the probability that the daisy allele successfully cuts and copies the engineered allele in a heterozygote at the other locus.

To better understand this equation, note that we assume all three loci undergo independent segregation, so we multiply the inheritance probabilities at each locus to obtain the overall probability. The daisy locus undergoes simple Mendelian inheritance, as is reflected in the term before the product. The other two cargo loci are identically influenced by the first locus, so they are put together in the product. Inside the product, the first term produces Mendelian inheritance if there is no daisy allele, while the second term introduces drive if there is a daisy allele.

Let c_i be the fitness cost of having at least one engineered allele at locus *i.* The fitness of genotype $\alpha\beta$ is given by:

$$f_{\alpha\beta} = \prod_{i=1}^{3} 1 - c_i\left(1 - \delta_{\alpha_i 0}\delta_{\beta_i 0}\right).$$

Finally, the rate of change of the genotype $[a, b]$ is given by the following set of equations:

$$\dot{x}_{ab} = v_{ab} \sum_{\alpha} g_\alpha \sum_{\beta} g_\beta \prod_i \Delta_{a_i b_i}^{\alpha_i \beta_i} - \psi x_{ab}$$

There is one equation of this type for every distinct genotype.

The idea behind these equations is that the birth rate of a genotype is proportional to the frequencies of the two haplotypes that make up the genotype, given that they can produce a viable individual. Whether a pair of haplotypes comprise a genotype is determined by $\Delta_{a_i b_i}^{\alpha_i \beta_i} = \delta_{a_i b_i}\delta_{\alpha_i a_i}\delta_{\beta_i b_i} + \left(1 - \delta_{a_i b_i}\right)\left(1 - \delta_{\alpha_i \beta_i}\right)$. The first term is 1 when both alleles at position *i* in the genotype are identical and when the alleles in both haplotypes at position *i* are correct, otherwise 0; the second term is 1 when the genotype is heterozygous at position *i* and the two haplotype alleles at that position are different, otherwise 0. To produce the quorum effect, only individuals with two functional copies each of the *A* and *B* genes are viable.

$v_{ab} = \delta_{a_2b_2}\delta_{a_3b_3}\delta_{a_2a_3} + \left(1 - \delta_{a_2b_2}\right)\left(1 - \delta_{a_3b_3}\right)$ is an indicator variable which is 1 if the genotype $[a, b]$ is viable and 0 otherwise. It sets the birth rate at zero if the genotype $[a, b]$ is nonviable.

The quantity ψ simply enforces the constraint that the population size remains constant. This is given by:

$$\psi = \sum_{ab} v_{ab} \sum_{\alpha} g_{\alpha} \sum_{\beta} g_{\beta} \prod_{i} \Delta_{a_i b_i}^{\alpha_i \beta_i}$$

The outer sum is over all distinct genotypes.

Daisy Quorum Gene Drive in Two Populations

We now extend our model from the previous section to include a basic spatial component: two populations connected by gene flow.

Consider a system with a nonzero rate of migration of individuals between the "target" population (where the initial release of engineered individuals occurs) and the non-target population. Let X_{ab} (with an uppercase *X*) be the frequency of individuals with genotype [*a,b*] in the target population, and Y_{ab} be the same for individuals in the non-target population. Similarly, let $G_a^{(T)}$ and $G_a^{(N)}$ be the corresponding gamete production rates. These are given by:

$$G_a^{(T)} = \sum_{\alpha\beta} X_{\alpha\beta} f_{\alpha\beta} P_{\alpha\beta,a}$$

$$G_a^{(N)} = \sum_{\alpha\beta} Y_{\alpha\beta} f_{\alpha\beta} P_{\alpha\beta,a}$$

Next, let the target population be a fraction *R* of the total population, with the non-target population as a fraction 1-*R*. To impose constant size constraints, assume that the total number of individuals migrating in each direction is equal, and the rate of migration is *r*. The selection dynamics for daisy quorum gene drive in two populations connected by gene flow are modeled by:

$$\dot{X}_{ab} = v_{ab} \sum_{\alpha} G_{\alpha}^{(T)} \sum_{\beta} G_{\beta}^{(T)} \prod_{i} \Delta_{a_i b_i}^{\alpha_i \beta_i} + \frac{r}{R}(Y_{ab} - X_{ab}) - \psi^{(T)} X_{ab}$$

$$\dot{Y}_{ab} = v_{ab} \sum_{\alpha} G_{\alpha}^{(N)} \sum_{\beta} G_{\beta}^{(N)} \prod_{i} \Delta_{a_i b_i}^{\alpha_i \beta_i} + \frac{r}{1 - R}(X_{ab} - Y_{ab}) - \psi^{(N)} Y_{ab}$$

To impose density constraints such that the frequencies sum to 1 in both populations, set:

$$\psi^{(T)} = \sum_{ab} v_{ab} \sum_{\alpha} G_{\alpha}^{(T)} \sum_{\beta} G_{\beta}^{(T)} \prod_{i} \Delta_{a_i b_i}^{\alpha_i \beta_i}$$

$$\psi^{(N)} = \sum_{ab} v_{ab} \sum_{\alpha} G_{\alpha}^{(N)} \sum_{\beta} G_{\beta}^{(N)} \prod_{i} \Delta_{a_i b_i}^{\alpha_i \beta_i}$$

Results

Daisy Quorum Gene Drive in One Population

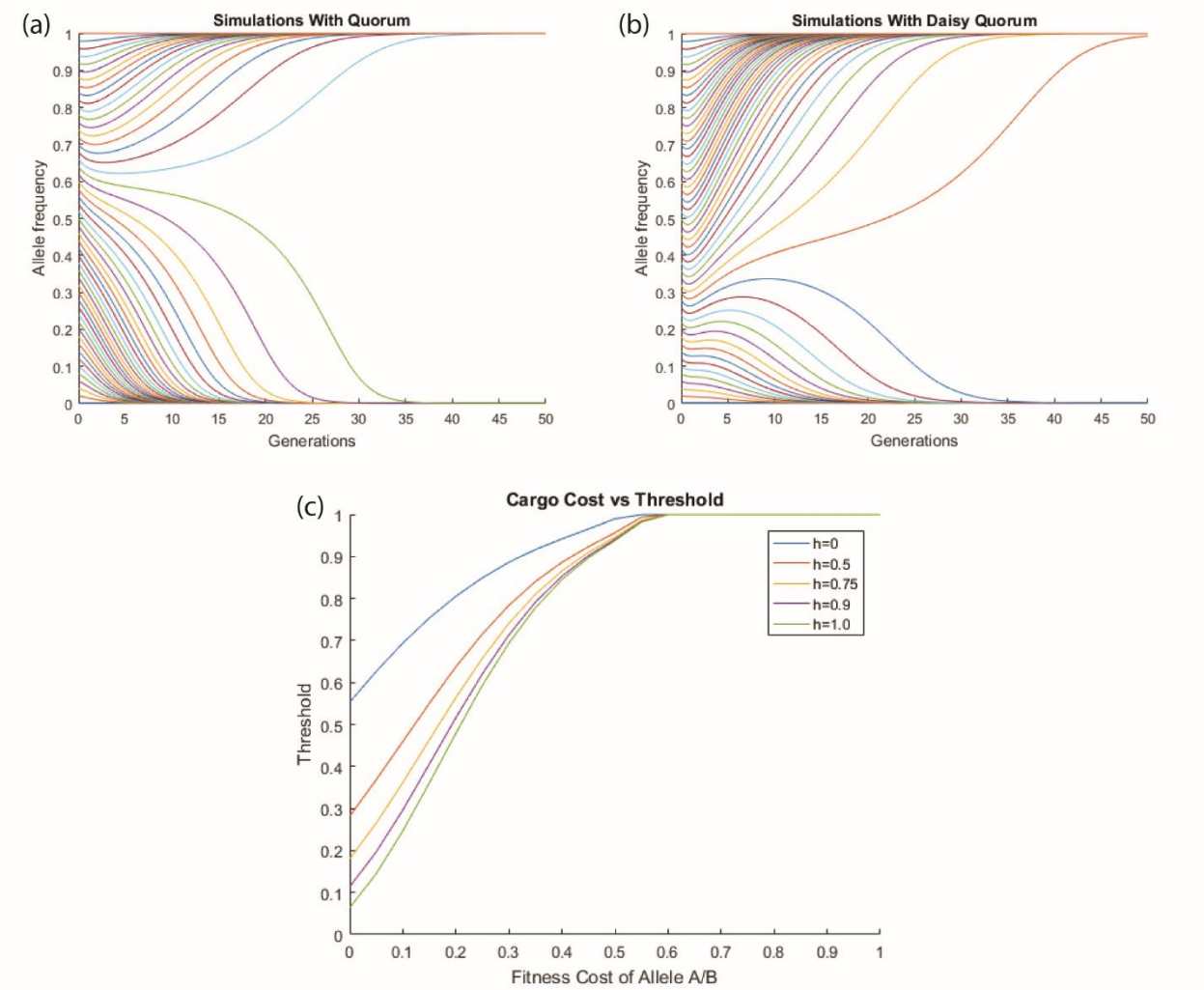

Figure 1. Thresholds for single-population quorum systems with and without daisy drive. Simulations were run with varying initial engineered individual frequencies to determine the threshold necessary for the engineered alleles to fix in the population. (a-b) Sample simulations (with the fitness cost of cargo alleles A and B set at 0.1) illustrate how releasing below the threshold frequency results in the engineered alleles will go extinct, while releasing above the threshold frequency results in fixation. (a) To create a quorum system without daisy drive using our model, we set the daisy allele C fitness cost $c_1 = 0$ and homing efficiency H = 0. The threshold engineered individual release frequency is 0.64. (b) To add the daisy drive element, we set the daisy allele C fitness cost $c_1 = 0.1$ and homing efficiency H = 0.9. The threshold engineered individual release frequency is 0.28. (c) Calculating thresholds by running simulations with varying homing efficiencies and fitness costs for cargo alleles A and B demonstrates that as the probability of drive increases (as measured by homing efficiency H), the threshold necessary for spread decreases.

Daisy Quorum Gene Drive in Two Populations

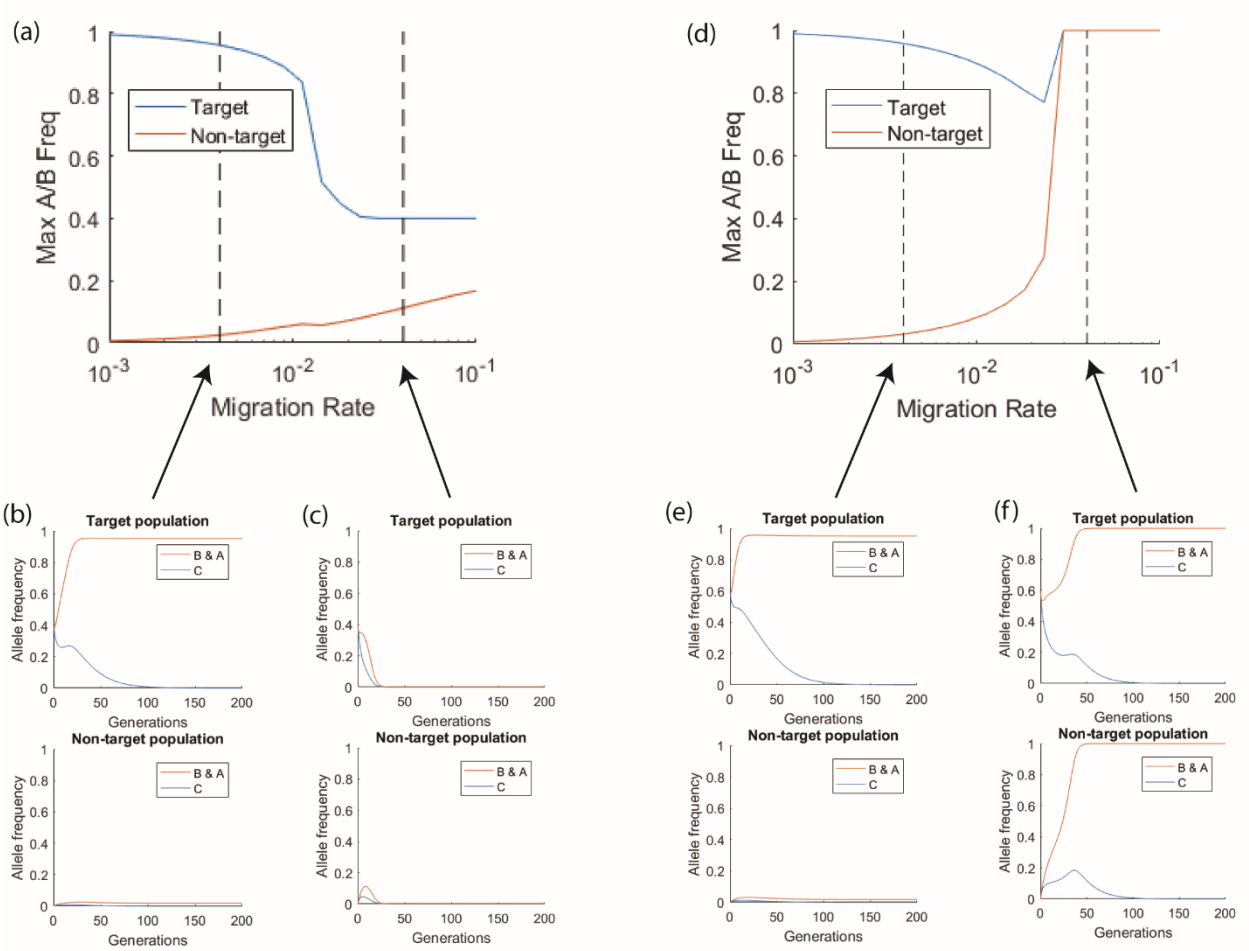

Figure 2. Containment of engineered individuals in the target population can be achieved with a low migration rate. If the migration rate is too high, then the engineered alleles will either fixate or go extinct in both populations, depending on the release frequency. The size of the target population matches the size of the non-target population, and homing efficiency $\mathrm{H} = 0.9$. The engineered individual frequency is set at 0.4 for (a-c), and it is set at 0.6 for (d-f). (a, d) As migration rate increases, the maximum spread of engineered cargo alleles generally decreases in the target population, but increases in the non-target population. The plateau at 0.4 for the target population is due to the release frequency being set at 0.4, showing that the engineered allele frequency only decreases over time. (d) At migration rate $\mathrm{r} = 4\mathrm{e} - 3$, for both (b) and (e), the engineered cargo alleles fix in the target population, but go extinct in the non-target population, which is the desired behavior for containment. At migration rate $\mathrm{r} = 4\mathrm{e} - 2$, (c) the engineered cargo alleles go extinct in both target and non-target populations with a lower release frequency, (f) but the engineered cargo alleles fix in both populations with a higher release frequency. This is what creates the spike in maximum spread of engineered alleles in the target population with higher release frequency in (d).

Future Directions

As originally proposed, we succeeded in confirming a quorum system produces a threshold effect and showed that introducing a daisy element decreases the threshold. We also demonstrated that containment of engineered alleles in the target population is possible when the migration rate between populations is low.

In the future, we'd like to investigate the time required for engineered alleles to spread. If the release frequency is above the threshold, how long will it take for the engineered alleles to fix? If the release frequency is below the threshold, how long will it take for the engineered alleles to die out?

We'd also like to investigate how the dynamics change as the ratios of target population size to non-target population size vary. Does a smaller target population size increase or decrease the threshold and migration rate necessary to have the engineered allele reach fixation in the target population while going extinct in the non-target population?

Acknowledgments

Charleston Noble, thank you so much for supervising me this semester!

References

1. Esvelt, K. M., Smidler, A. L., Catteruccia, F. & Church, G. M. Concerning RNA-guided gene drives for the alteration of wild populations. Elife 3, e03401 (2014).
2. Champer, J., Buchman, A. & Akbari, O. S. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. Nat. Rev. Genet. 17, 146–159 (2016).
3. Noble, C., Adlam, B., Church, G. M., Esvelt, K. M. & Nowak, M. A. Current CRISPR gene drive systems are likely to be highly invasive in wild populations. bioRxiv 219022 (2017). doi:10.1101/219022
4. Min, J., Noble, C., Najjar, D. & Esvelt, K. Daisy quorum drives for the genetic restoration of wild populations. bioRxiv (2017). doi:10.1101/115618
5. Noble, C. et al. Daisy-chain gene drives for the alteration of local populations. bioRxiv (2016). doi:10.1101/057307