Spatial gene drives and pushed genetic waves

Hidenori Tanaka, Howard A. Stone, and David Nelson

Mobolaji Williams — Shakhnovich Journal Club— Nov 10, 2017

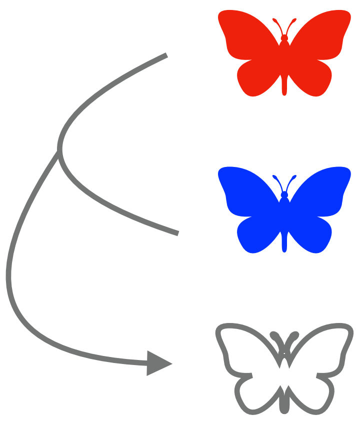
Mendelian Genetics 101

Question #1:

Microbiologists have managed to genetically engineer a mutant mosquito which does not propagate a certain disease.

Say that this mutant mosquito is AA for a particular trait and the wild type is aa for this trait. The mutant and the wild type mate and have an offspring.

What is the probability that said offspring will propagate the desired A allele into the next generation?

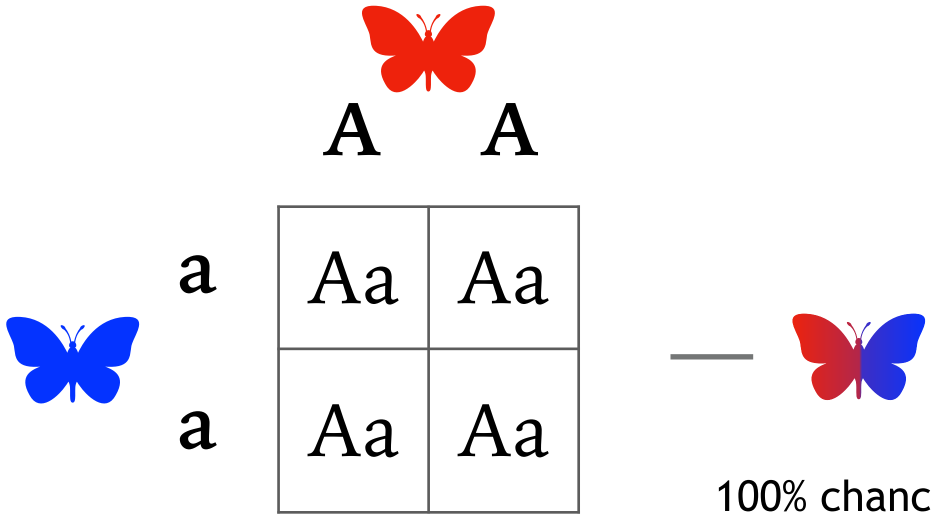


: Mutant type, *AA* genotype

: Wild type, *aa* genotype

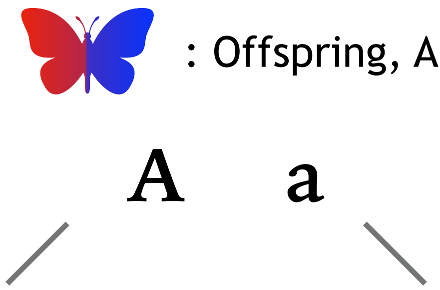
: Offspring, ??

Solution: Setup an inheritance matrix



	A	A
a	Aa	Aa
a	Aa	Aa

100% chance of Aa



: Offspring, Aa genotype

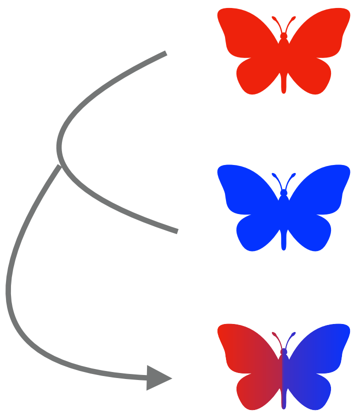
A a

50% chance of propagating A

50% chance of propagating a

There is a 50% chance that the offspring will propagate the allele A to the next generation.

Mendelian Genetics 101



: Mutant type, *AA* genotype

: Wild type, *aa* genotype

: Offspring, *Aa* genotype

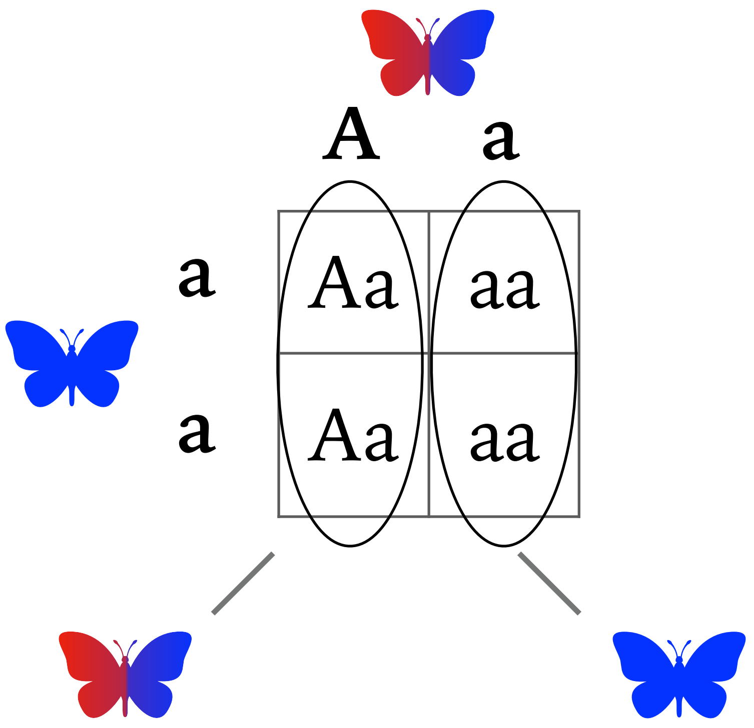
* Mixed colors refers to allele mixing, NOT phenotype mixing.

Question #2:

This mixed allele *Aa* offspring is placed in a population completely composed of Wild type *aa* organisms.

In its second generation, what proportion of Aa's offspring would have the allele A?

Solution: Setup an inheritance matrix

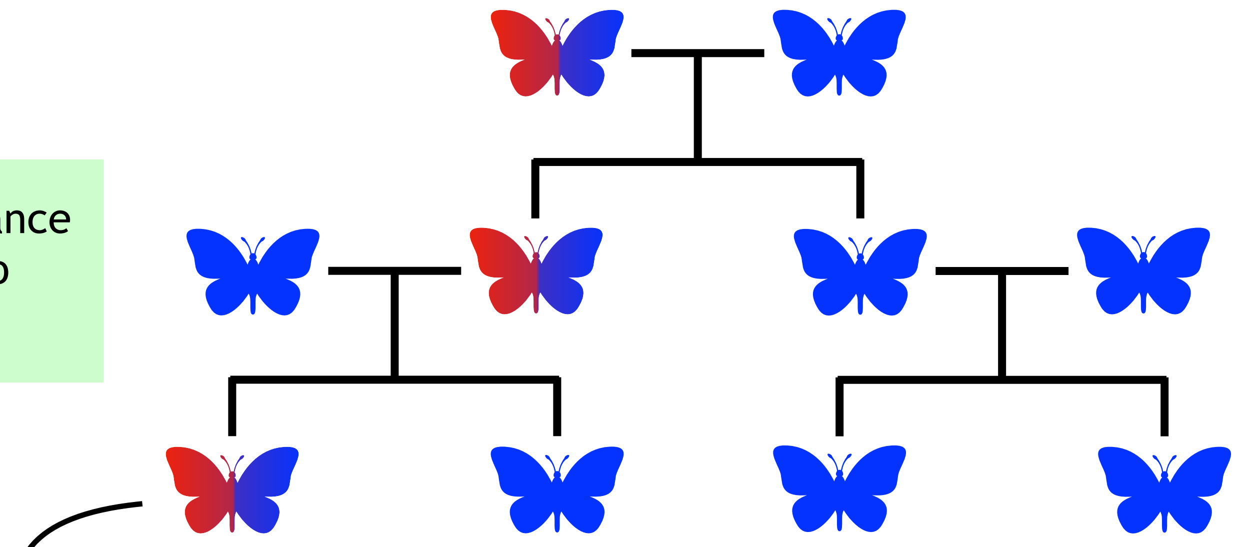


50% chance of Aa

50% chance of aa

There is a 50% chance that Aa passes A to offspring.

Mendelian Inheritance



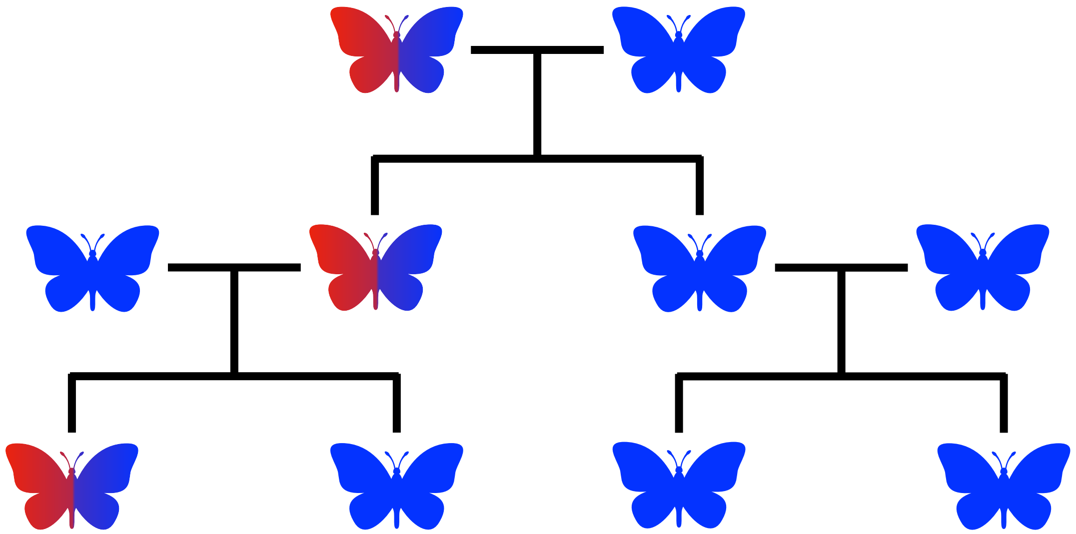
By the second generation, only %25 of the original Aa's offspring have the A allele

Why do we need Gene Drives?

A
Desired mutant allele

a
Wild allele

Mendelian Inheritance



By Aa’s second generation, only %25 of the original Aa’s offspring have the A allele

There’s a problem:

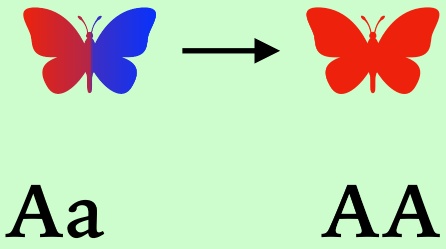
Our desired mutation does not propagate fast enough through the population.

Question: How can we get our desired mutation to comprise a larger proportion of the progeny?

Answer: Gene Drives!

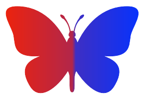
(Definition) Gene drives are genes which code for their own duplication in the partner chromosome.

(Functionally) Gene drives convert heterozygous genotypes into homozygous genotypes. (This occurs between generation cycles)



How do Gene Drives work?

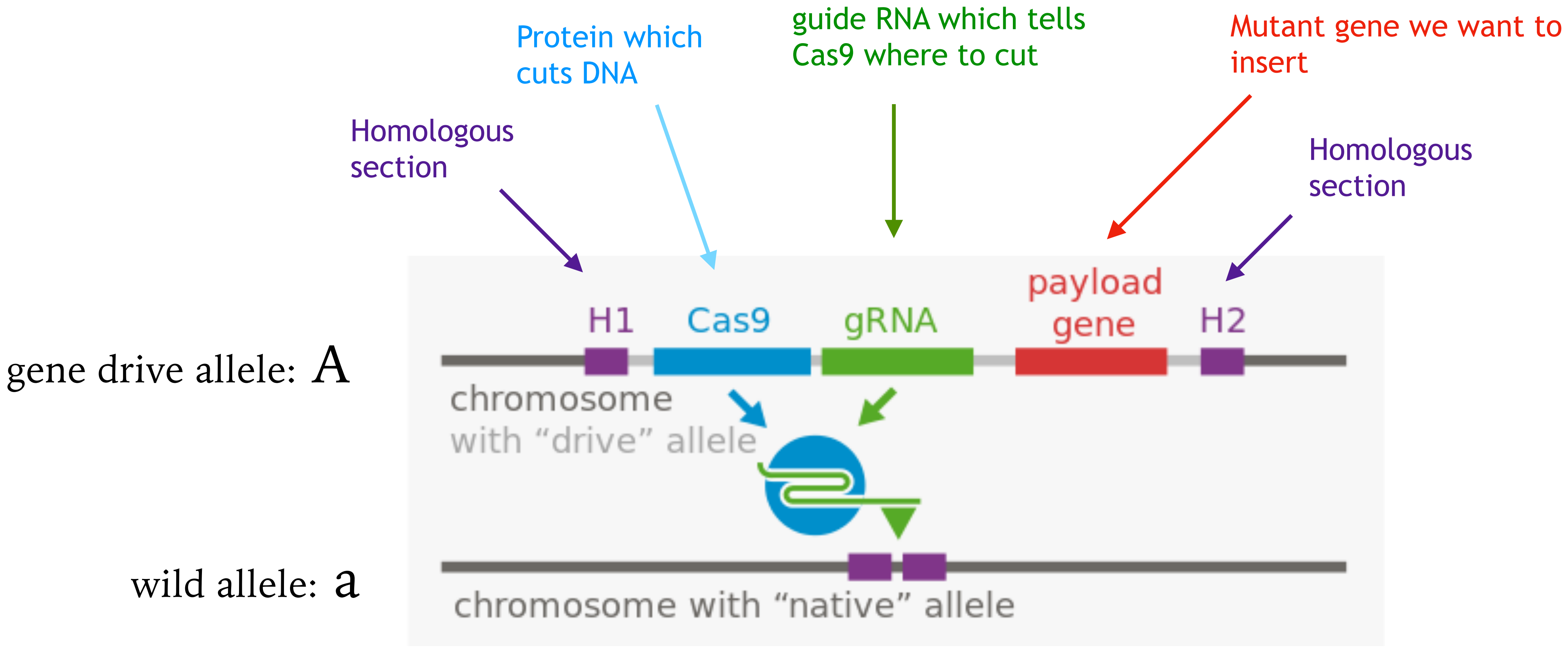
(Definition) Gene drives are genes which code for their own duplication in the partner chromosome.

Say we begin with the heterozygous mutant: 

A a

Mutant allele
(as a gene drive)

Wild allele



Protein which cuts DNA

guide RNA which tells Cas9 where to cut

Mutant gene we want to insert

Homologous section

Homologous section

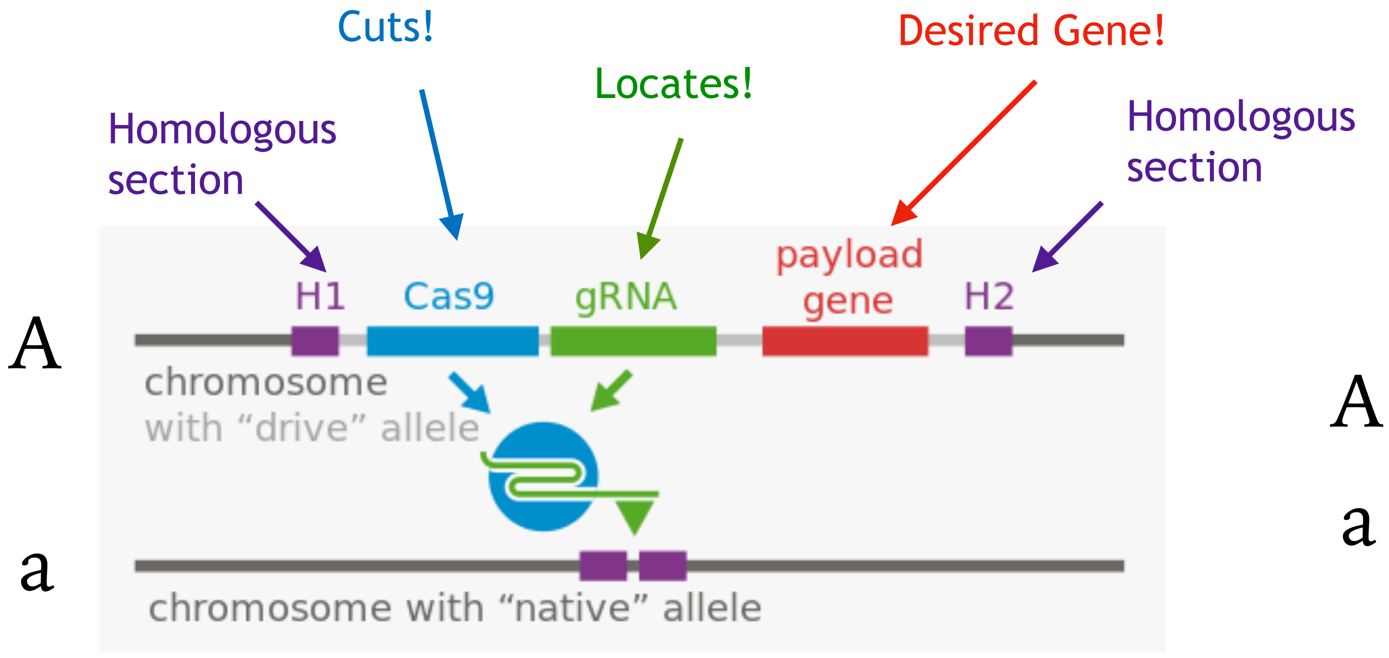
gene drive allele: A

wild allele: a

(By Thomas Julou - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=49101806)

How do Gene Drives work?

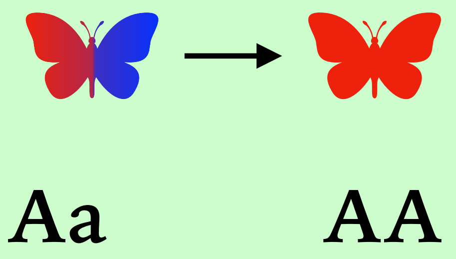
(Definition) Gene drives are genes which code for their own duplication in the partner chromosome.



#1) gRNA tells Cas9 where to cut partner gene

#2) Cut is repaired by replicating* gene drive in partner chromosome

NET RESULT:



* Replication process is prone to errors, so that it seldom occurs at 100% efficiency. (Paper assumes 100% efficiency)

(By Thomas Julou - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=49101806)

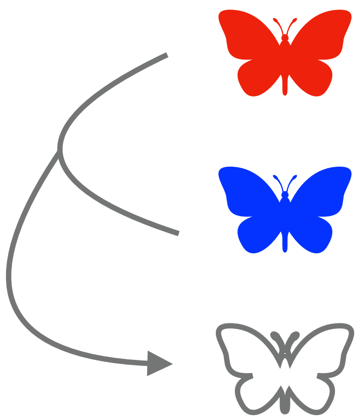
“Super Mendelian” Genetics 101

Question #1:

Microbiologists have managed to genetically engineer a mutant mosquito which does not propagate a certain disease.

Say that this mutant mosquito is AA (with a **GENE DRIVE**) for a particular trait and the wild type is aa for this trait. The mutant and the wild type mate and have an offspring.

What is the probability that said offspring will propagate the desired A allele into the next generation?

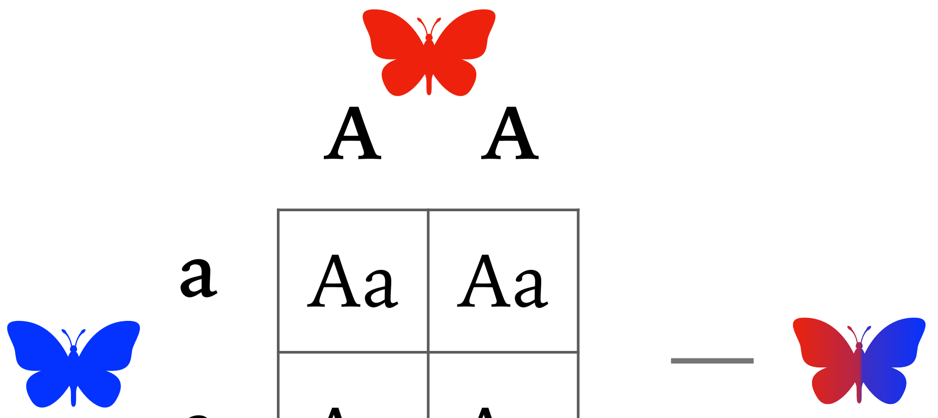


: Mutant type, *AA* genotype

: Wild type, *aa* genotype

: Offspring, ??

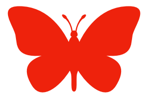
Solution: Setup an inheritance matrix



	A	A
a	Aa	Aa
a	Aa	Aa

All offspring are Aa

Gene drive converts Aa to AA



: Offspring, AA genotype

A A

There is a 100% chance that the offspring will propagate the allele A to the next generation.

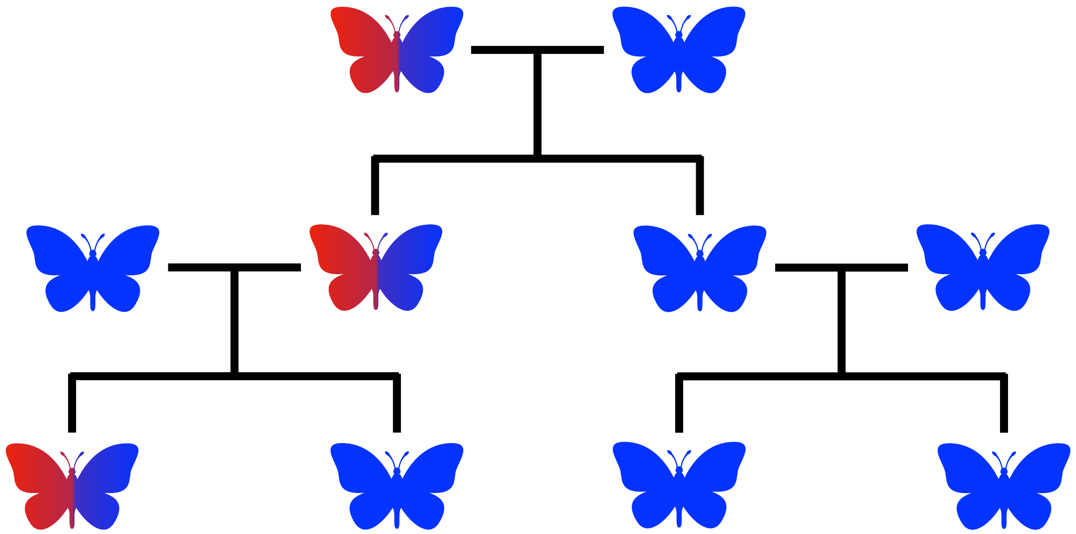
“Super Mendelian” Genetics 101

Question: How can we get our desired mutation to comprise a larger proportion of the progeny?

A
Desired mutant allele

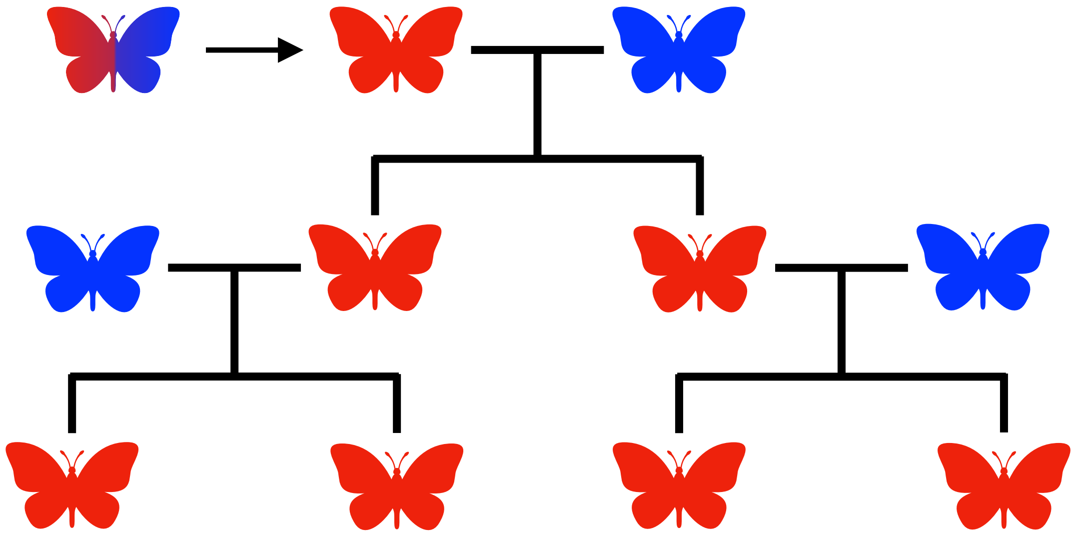
a
Wild allele

Mendelian Inheritance



By the second generation, only %25 of the original Aa’s offspring have the A allele

“Super Mendelian”/Gene Drive Inheritance



By the second generation, 100% of the original Aa’s offspring have the A allele.

Problem: Gene drive overcorrects Mendelian genetics!

Previously (Mendelian Genetics):

Engineered mutation does not propagate fast enough through the population.

Now (Super Mendelian Genetics):

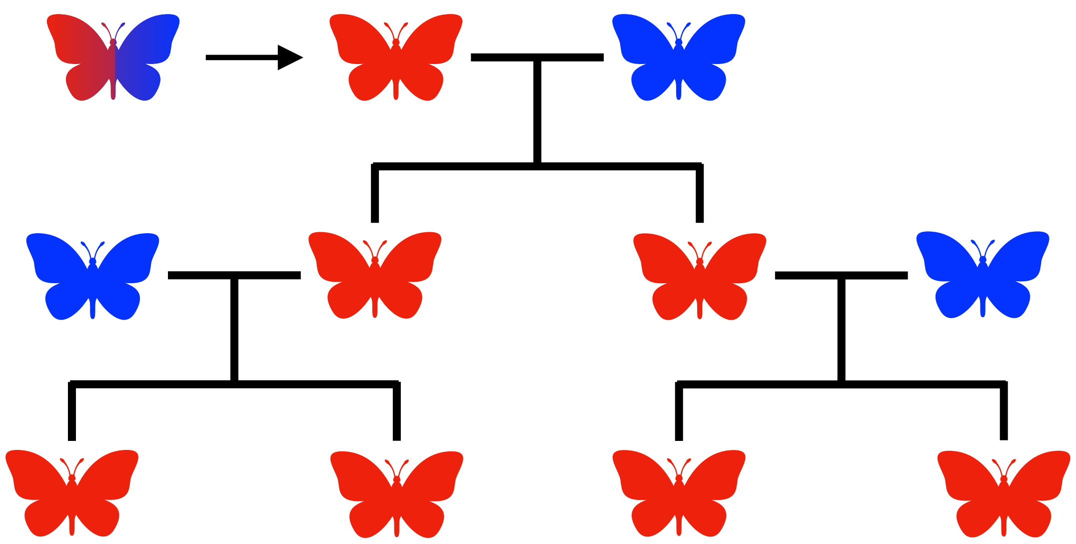
Engineered mutation (with gene drive) completely overtakes the population.

“Super Mendelian” Genetics 101

A
Desired mutant allele

a
Wild allele

“Super Mendelian”/Gene Drive Inheritance



By the second generation, 100% of the original Aa’s offspring have the A allele.

AND 0% have the a allele

Paper’s Main Question:

How can we control whether a gene-drive mutation overtakes a population?

Partial Answer:

The gene drive mutant has a selective disadvantage relative to the wild type.

$$w_{\mathrm{drive}} = w_0(1-s)$$

Main Question (Rephrased): For what selection coefficients can an organism with a gene drive be controlled so that it neither overtakes nor is eliminated from a population?

Genetics through Hardy-Weinberg Equilibrium

Paper's Main Question: For what selection coefficients can an organism with a gene drive be controlled so that it neither overtakes nor is eliminated from a population?

Main Strategy: Use Hardy-Weinberg Principles to derive an equation for allele frequency as a function of time.

$$w_{\text{drive}} = w_0(1-s)$$

What is s for a controllable gene drive?

q : frequency of gene drive allele

$$\frac{dq}{dt} + U'(q,s) = 0$$

Hardy-Weinberg (HW) Principle: The allele frequencies in a population (without selection and mutations) remains constant throughout time.

Example 1: HW with NO SELECTION and NO GENE DRIVE

q: frequency of "A" (mutant allele) at time t

p: frequency of "a" (wild allele) at time t

*$p+q=1$ (frequencies sum to 1)

Genotype	Frequency	Fitness
AA	q^2	w_0
Aa	$2pq$	w_0
aa	p^2	w_0

How does the allele frequency of A change for an incremental time step?

$$\begin{aligned} q(t+\Delta t) &= \frac{\text{Mean fitness for allele A}}{\text{Mean fitness overall}} \\ &= \frac{w_0 q^2 + \frac{1}{2} w_0 2pq}{w_0 q^2 + w_0 2pq + w_0 q^2} \\ &= q^2 + qp \\ &= q \end{aligned}$$

As we expect, the frequency of allele A remains the same for all time changes.

Genetics through Hardy-Weinberg Equilibrium

Example 2: HW with SELECTION and NO GENE DRIVE

q : frequency of "A" (mutant allele) at time t

p: frequency of "a" (wild allele) at time t

$^*p + q = 1$ (frequencies sum to 1)

Genotype	Frequency	Fitness
AA	q^2	$w_0(1-s)$
Aa	$2pq$	w_0
aa	p^2	w_0

How does the allele frequency of A change for an incremental time step?

$$
\begin{aligned}
q(t+\Delta t) &= \frac{\text{Mean fitness for allele A}}{\text{Mean fitness overall}} \\
&= \frac{w_0(1-s)q^2 + \frac{1}{2}w_0 2pq}{w_0(1-s)q^2 + w_0 2pq + w_0 q^2} \\
&= \frac{q(1-sq)}{1-sq^2}
\end{aligned}
$$

We find that *q(t+Δt)* < *q*. Thus when the mutant genotype is selected against, its frequency decreases

---

Example3: HW with SELECTION and GENE DRIVE (case in the paper)

q : frequency of "A" (mutant allele) at time t

p: frequency of "a" (wild allele) at time t

$^*p + q = 1$ (frequencies sum to 1)

How does the allele frequency of A change for an incremental time step?

Genotype	Frequency	Fitness
AA	q^2	$w_0(1-s)$
Aa	$2pq$	w_0
aa	p^2	w_0

Action of Gene Drive

$$\mathrm{Aa} \longrightarrow \mathrm{AA}$$

Genotype	Frequency	Fitness
AA	$q^2 + 2pq$	$w_0(1-s)$
aa	p^2	w_0

Genetics through Hardy-Weinberg Equilibrium

Example3: HW with SELECTION and GENE DRIVE (case in the paper)

q: frequency of "A" (mutant allele) at time t

p: frequency of "a" (wild allele) at time t

$^*p + q = 1$ (frequencies sum to 1)

How does the allele frequency of A change for an incremental time step?

Genotype	Frequency	Fitness
AA	q^2	$w_0(1-s)$
Aa	$2pq$	w_0
Aa	p^2	w_0

Action of Gene Drive

$$\text{Aa} \longrightarrow \text{AA}$$

Genotype	Frequency	Fitness
AA	$q^2 + 2pq$	$w_0(1-s)$
aa	p^2	w_0

$$q(t+\Delta t) = \frac{\text{Mean fitness for allele A}}{\text{Mean fitness overall}}$$

$$= \frac{w_0(1-s)(q^2+2pq)}{p^2 + w_0(1-s)(q^2+2pq)}$$

$$= \frac{(1-s)q(q+2p)}{1-sq(2-q)}$$

It is unclear how *q(t+Δt)* compares to *q*. We require a more involved analysis.

Deriving differential equation

Define:

$$\lim_{\Delta t \to 0} q(t+\Delta t) - q \equiv \tau_g \frac{dq}{dt}$$

We then obtain

τ_g: generation time

$$\boxed{\tau_g \frac{dq}{dt} = \frac{sq(1-q)(q-q^*)}{1-sq(2-q)} \qquad \text{where } q^* \equiv \frac{2s-1}{s}}$$

Genetics through Hardy-Weinberg Equilibrium

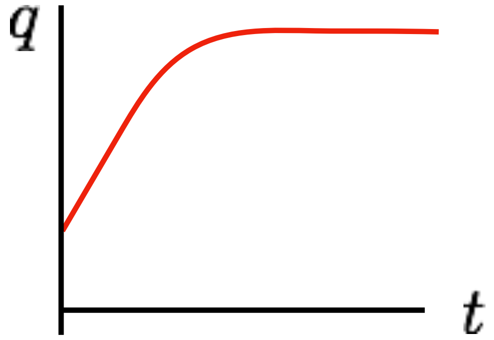
HW with SELECTION and GENE DRIVE (case in the paper)

$$\boxed{\tau_g \frac{dq}{dt} = \frac{sq(1-q)(q-q^*)}{1-sq(2-q)} \qquad \text{where } q^* \equiv \frac{2s-1}{s}}$$

This equation determines how the mutant (gene drive) frequency evolves under selection

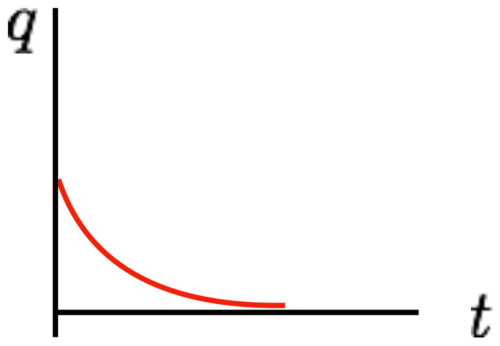
Limiting Cases: No Selection and Full Selection

– No Selection; $s = 0$
(Mutant has wild type fitness) : $\tau_g \frac{dq}{dt} = q(1-q)$



The mutant (gene drive) allele completely overtakes the population

– Full Selection; $s = 1$
(Mutant has no fitness) : $\tau_g \frac{dq}{dt} = -q$



The mutant (gene drive) allele completely dies out

Reality is somewhere in between.

How do we analyze the long time behavior of the mutant allele for intermediate values of *s*?

Potential Energy Landscape for Long-Time Behavior

How do we analyze the long-time behavior of q(t) for arbitrary selection coefficient *s*?

$$\tau_g \frac{dq}{dt} = \frac{sq(1-q)(q-q^*)}{1-sq(2-q)}$$

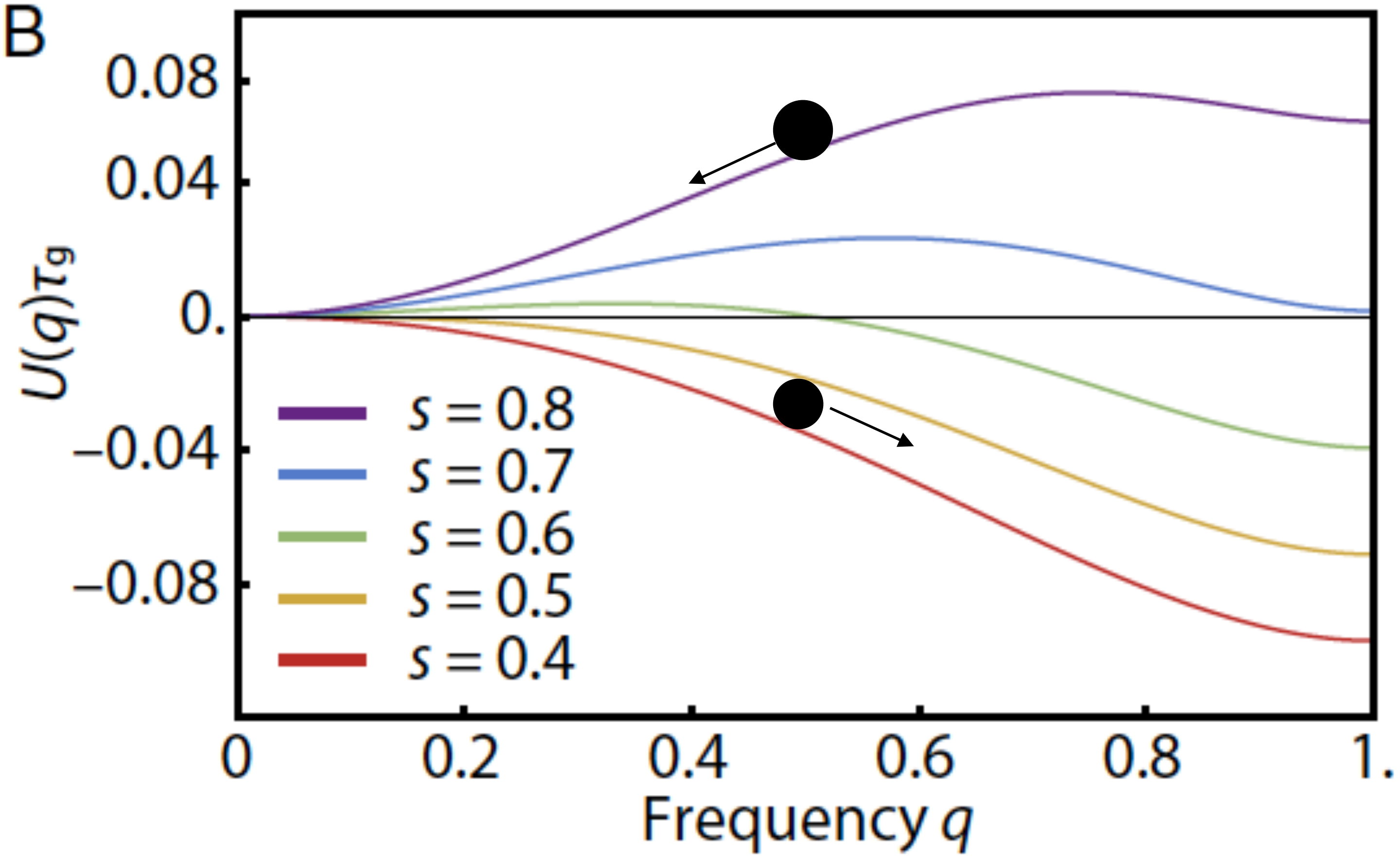
$$\text{where } q^* \equiv \frac{2s-1}{s}$$

Introduce a Potential Function

$$\frac{dq}{dt} = -U'(q,s)$$

$$U(q,s) = -\frac{1}{\tau_g}\int_0^q dq' \, \frac{sq(1-q)(q-q^*)}{1-sq(2-q)}$$

Potential Landscape



Long Time Behavior:
For a given initial q, we can find the long time behavior by letting q "roll" down the potential landscape

The system evolves to either *q* = 0 or *q* =1.

Say $q(t=0) = 0.5$

— High selection coefficient (s = 0.8) $q(t \to \infty) = 0$

— Low selection coefficient (s = 0.4) $q(t \to \infty) = 1$

Gene Drive Dynamics with Diffusion

How do the dynamics change when the alleles are allowed to diffuse through space?

Answer found by solving the 1D Diffusion equation

$$\tau_g \frac{\partial q}{\partial t} = D \frac{\partial^2 q}{\partial x^2} + \frac{sq(1-q)(q-q^*)}{1-sq(2-q)}$$

Qualitatively, the authors find three regimes

Explosion (s<0.5):

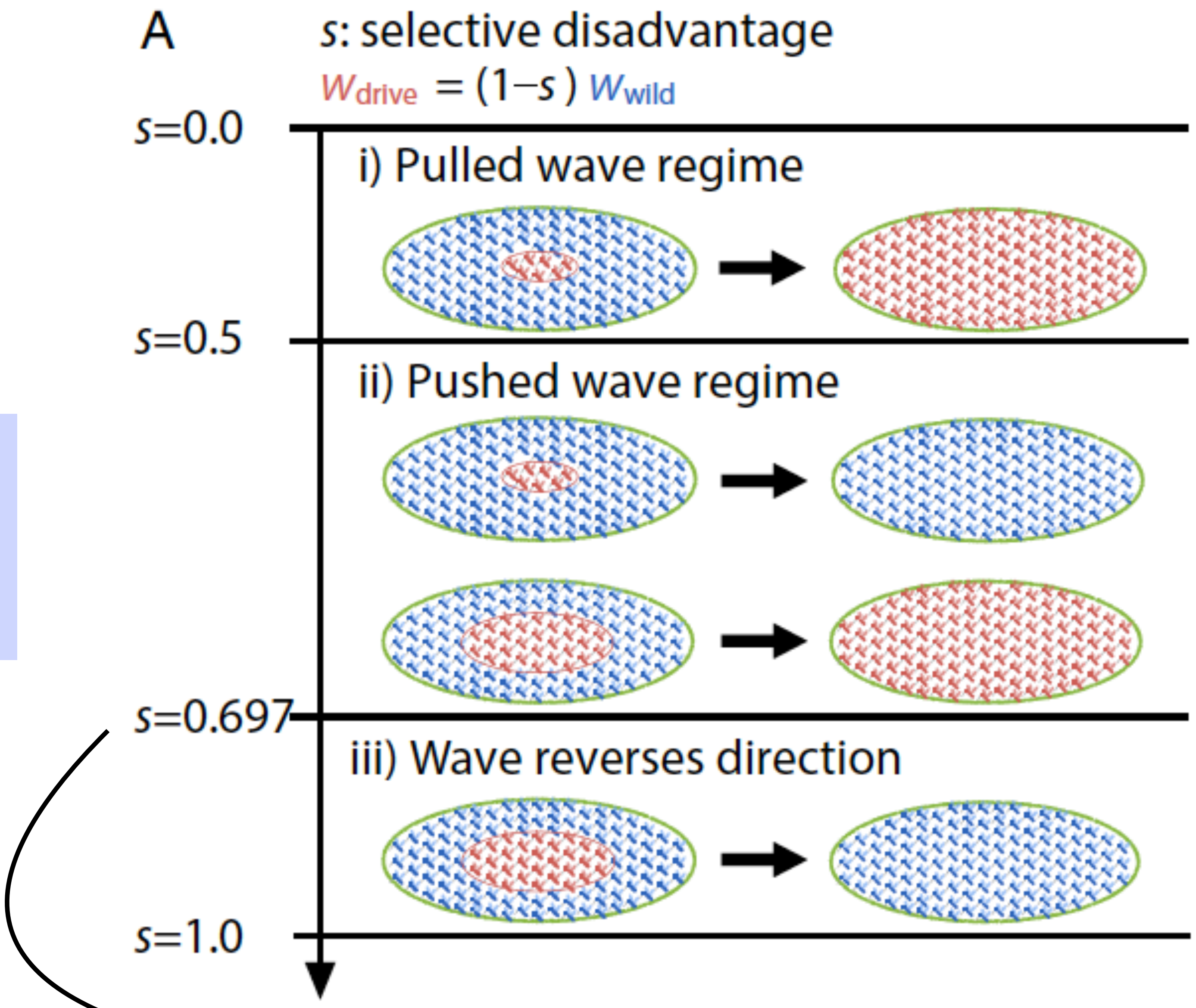
Gene drive allele
ALWAYS TAKES OVER

Tunable (0.5<s<0.697):

Gene drive allele only
takes over for certain conditions

Failure (s>0.697):

Gene drive allele
ALWAYS COLLAPSES



*Value of s where $U(0,s) = U(1,s)$.

Specific Solutions to Gene Drive Diffusion Equation

How does the nature of solutions to

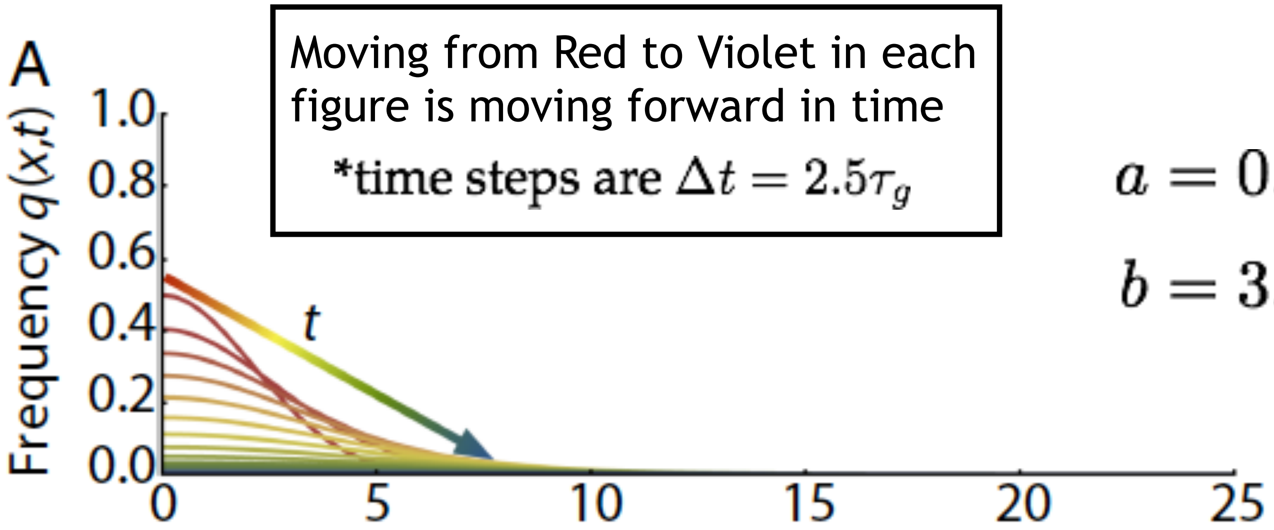
$$\tau_g \frac{\partial q}{\partial t} = D\frac{\partial^2 q}{\partial x^2} + \frac{sq(1-q)(q-q^*)}{1-sq(2-q)}$$

depend on solution parameters and selection coefficient?

Defining initial condition and selection coefficient:

$$q(x,0) = a\,e^{-(x/b)^2}\,;\, s = 0.58$$

*(Tunable Regime)
0.5<0.58<0.697



Moving from Red to Violet in each figure is moving forward in time

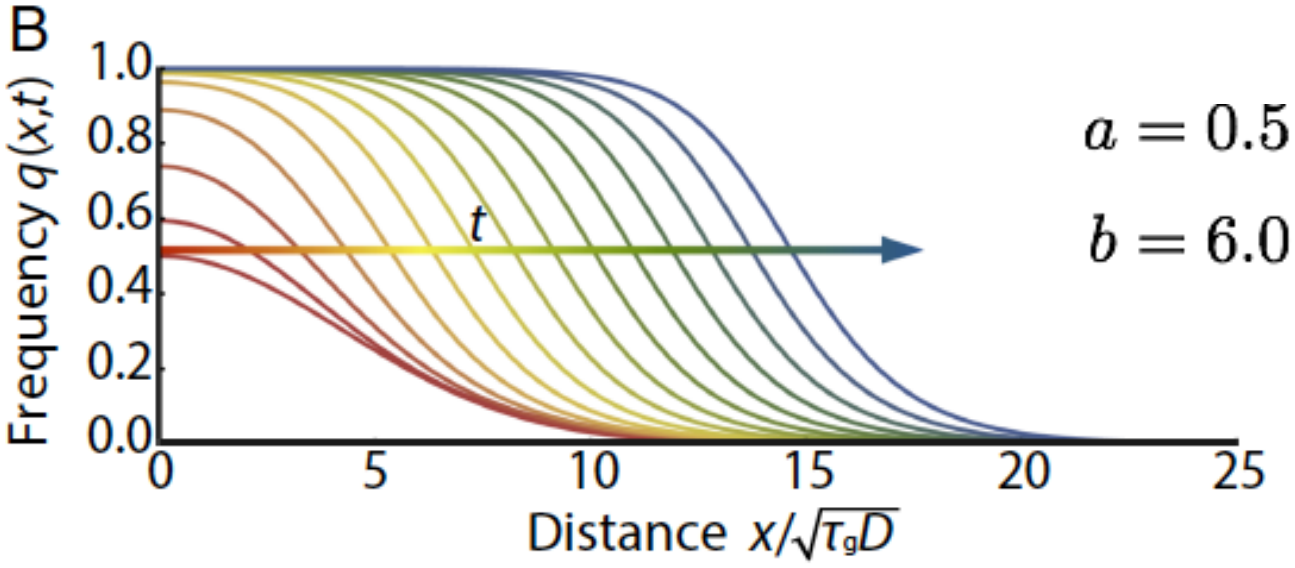
*time steps are $\Delta t = 2.5\tau_g$

$$a = 0.5$$

$$b = 3.0\sqrt{\tau_g D}$$

Narrow distribution:

Gene drive allele eventually collapses



$$a = 0.5$$

$$b = 6.0\sqrt{\tau_g D}$$

Broad distribution:

Gene drive allele eventually takes over.

Gene Drive Diffusion Equation with Selection Barriers

What is the nature of these solutions when we incorporate a **selection boundary**?

$$\tau_g \frac{\partial q}{\partial t} = D\frac{\partial^2 q}{\partial x^2} + \frac{sq(1-q)(q-q^*)}{1-sq(2-q)}$$

Defining initial condition and function parameters

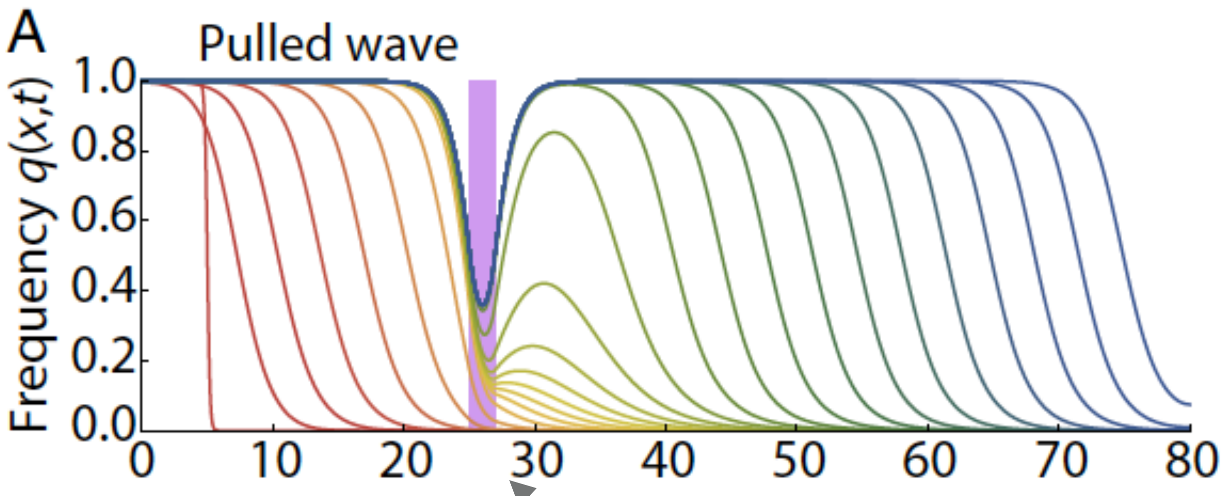
$$q(x,0) = \frac{q_0}{1+e^{10(x-x_0)/\sqrt{\tau_g D}}}$$

where

$$q_0 = 1.0$$

$$x_0 = 5.0\sqrt{\tau_g D}$$

$$s = 0.479 < 0.5 = s_{\min}$$



Moving from Red to Violet in each figure is moving forward in time

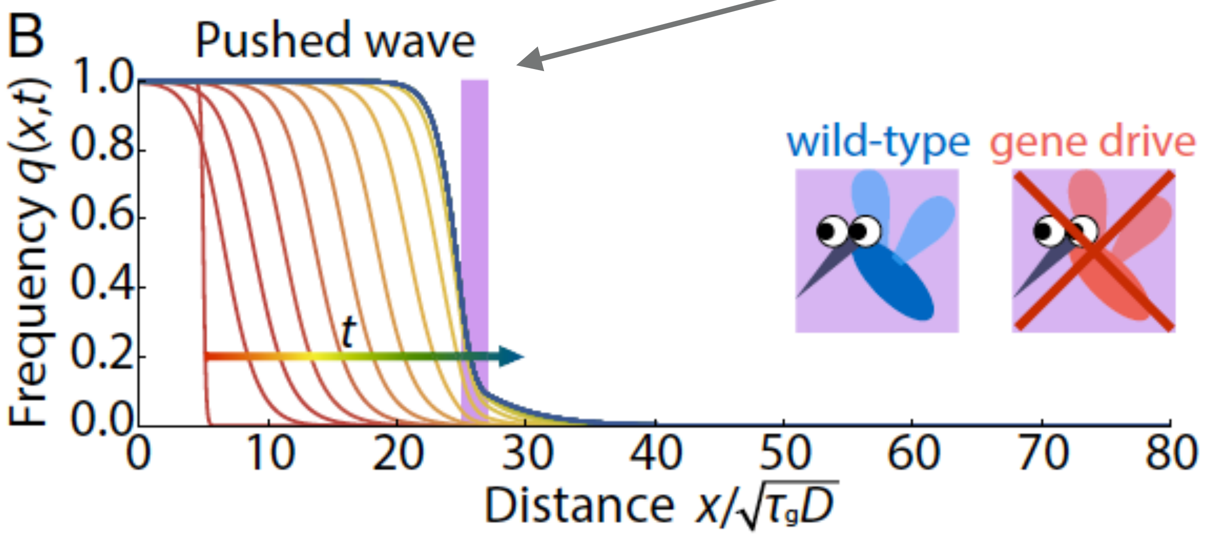
*time steps are $\Delta t = 2.5\tau_g$

Low (s<0.5) Selection:

Gene drive allele overcomes selection barrier

$$s_b = 0.958$$

$$s = 0.542 > 0.5 = s_{\min}$$



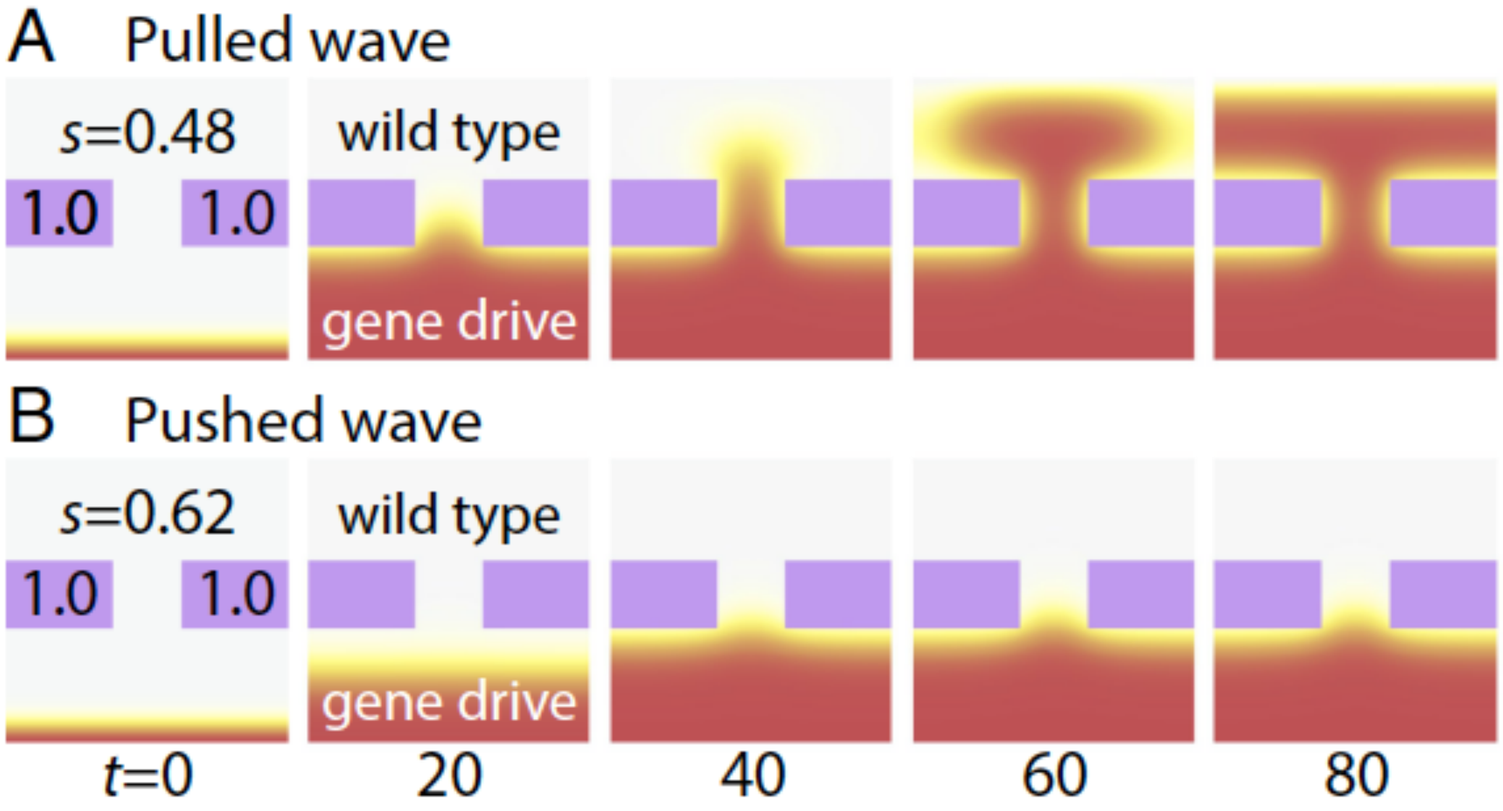
Tunable (0.5< s< 0.697) Selection:

Gene drive allele cannot get past selection barrier

Selection Barriers in Two-Dimensions

How do selection barriers affect the propagation of alleles for two dimensional systems?

$$\tau_g \frac{\partial q}{\partial t} = D\left(\frac{\partial^2 q}{\partial x^2} + \frac{\partial^2 q}{\partial y^2}\right) + \frac{sq(1-q)(q-q^*)}{1-sq(2-q)}$$



Low (s<0.5) selection:

Allele is able to diffuse through gap in the barrier.

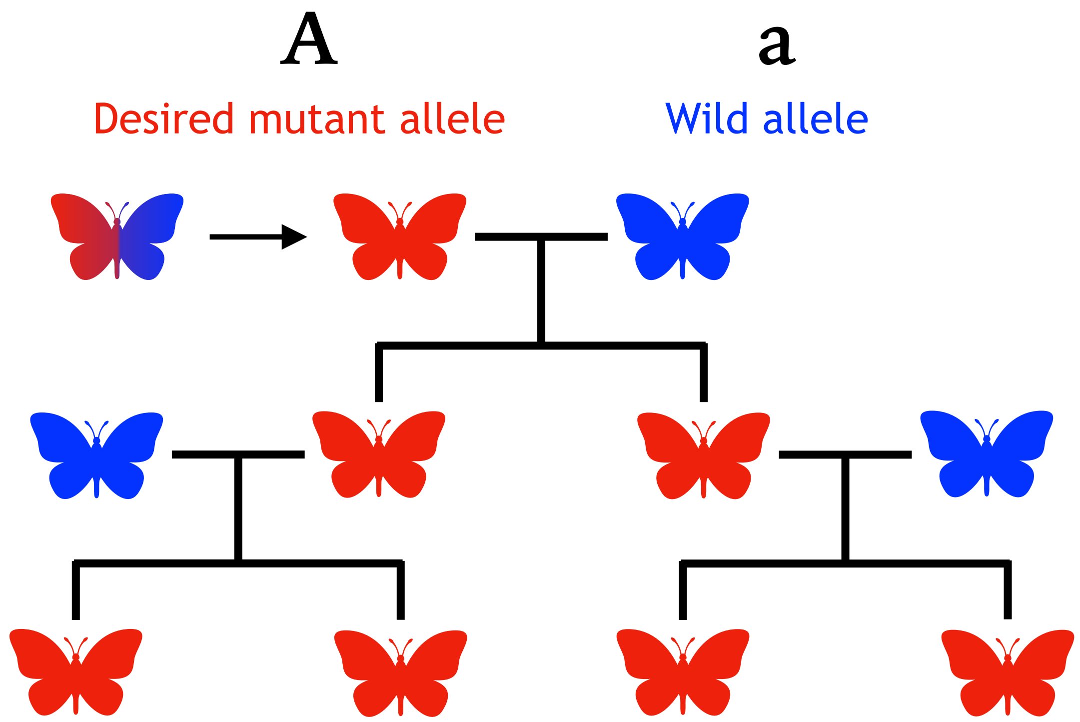
Tunable(0.5<s<0.697) selection:

Allele cannot diffuse through gap in the barrier

Even incomplete barriers in 2D can limit the spread of an allele

Concluding Summary

“Super Mendelian”/Gene Drive Inheritance



A Desired mutant allele

a Wild allele

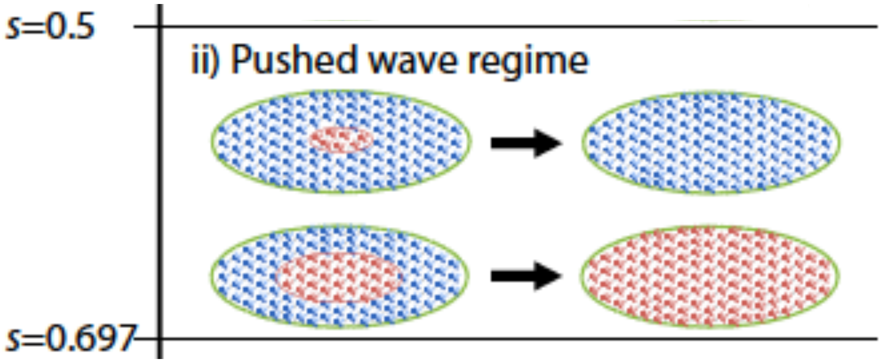
Paper’s Main Question:

How can we control whether a gene-drive mutation overtakes a population?

General Answer:

We can limit the spread of the gene-drive mutation by engineering its selection coefficient to be within a specific range of values

*For the model in the paper the range is 0.5 <s < 0.697



Some other questions

- How does inefficient gene-drive conversion change the results? (Answered in appendix)
- How would one specifically engineer a desired selection coefficient?

END