Strong Selection is Necessary for Evolution of Blindness in Cave Dwellers

Rachel S. Schwartz,[1] Alexandra L. Merry,[2] Megan M. Howell,[2] and Reed A. Cartwright[1,3,*]

[1]The Biodesign Institute

[2]Barrett, The Honors College

[3]School of Life Sciences

Arizona State University

Tempe, AZ, USA

[*]Corresponding author: cartwright@asu.edu

Running title: Blindness is Due to Strong Selection in Caves

Key words: Cavefish, Model, Drift

1 **Abstract**

2 Blindness has evolved repeatedly in cave-dwelling organisms, and investigating loss of sight presents an

3 opportunity to understand the operation of fundamental evolutionary processes, including drift, selection,

4 mutation, and migration. The observation of blind organisms has prompted many theories for their

5 blindness, including loss-by-disuse and selection against eye development when eyes are not used. Here

6 we have developed a model that shows just how strong selection must be for blind populations of a cave-

7 dwelling species to evolve. We used approximations to determine levels of selection that would result in

8 caves containing only sighted individuals, only blind individuals, or a stable population of both. We then

9 incorporated drift into the model using simulations. Based on our model, strong selection is necessary

10 for the evolution of blindness unless immigration rates are extremely low. Drift decreased the fixation of

11 blindness in populations, although for intermediate levels of migration the level of selection required to fix

12 blindness decreased substantially. We hypothesize that this degree of selection may be due to phototaxis

13 in sighted individuals, who move toward the light leaving only blind individuals in the cave.

Introduction

Blindness has evolved repeatedly across taxa in caves, creating nearly a thousand cave-dwelling species and many more populations (Culver et al., 2000; Dowling et al., 2002; Bradic et al., 2012; Coghill et al., 2014). However, many populations of blind individuals experience some level of immigration, which would be expected to prevent the fixation of blindness in a newly established population (Avise and Selander, 1972; Bradic et al., 2012; Coghill et al., 2014). Thus, blind cave-dwelling populations of typically sighted species pose an interesting challenge to our understanding of evolutionary biology. Namely, how does a fixed phenotype evolve from low frequency despite immigration?

Darwin suggested that eyes would be lost by "disuse" (Darwin, 1859). We now consider this hypothesis the "neutral mutation hypothesis" — random mutations can accumulate in eye related genes or regulatory regions when, as in caves, there is no purifying selection to eliminate them. However, the accumulation of mutations (mutation pressure) causing blindness would take a long time to result in fixation of blindness in populations on its own (Barr, 1968). Thus, it is genetic drift combined with mutation pressure that would lead to blindness (Kimura and King, 1979; Borowsky, 2015). Genetic drift is the increases the frequency of blindness alleles created by mutations: eyes become increasingly less functional and finally disappear completely (Wilkens, 1988).

There are a variety of cave dwelling vertebrates and invertebrates (referred to as cavernicoles, troglophiles, and troglobites). However, much of the work on the evolution of blindness has focused on the blind form of cavefishes, e.g. the Mexican tetra (*Astyanax mexicanus*) and Atlantic molly (*Poecilia mexicana*). For example, the hypothesis of relaxed selection is supported by the observation of a high number of mutations in cavefish putative eye genes (Hinaux et al., 2013; Protas et al., 2006; Gross et al., 2009). However, developmental evidence does not support this hypothesis: cavefish embryos begin eye development, but the eye disappears in larvae (Langecker et al., 1993; Jeffery et al., 2003). Random mutations should occur in genes controlling early eye development as well. Furthermore, although drift would often lead to loss of seeing individuals joining a population of blind individuals, this model depends on developing a high frequency of blindness in a cave population simply by drift in isolation.

Alternatively, the "adaptation hypothesis" suggests that there is a cost to an eye; thus, individuals without

2

41 eyes have greater fitness resulting in the eventual elimination of seeing individuals. This cost may either

42 come from the energy required to develop a complex structure or due to the vulnerability of the eye (Barr,

43 1968; Strickler et al., 2007; Jeffery, 2005; Protas et al., 2007; Niven, 2008; Niven and Laughlin, 2008; Moran

44 et al., 2015). Another hypothesis states that sight is not lost not due to the "cost" of development but

45 due to pleiotropic mutations selected for other traits. For example, in Mexican tetra increased expression

46 of Hedgehog (Hh; Jeffery, 2005) likely affects feeding structures, allowing better foraging in low light

47 conditions (Jeffery, 2005, 2001). However, increased Hh signaling inhibits pax6 expression, which results

48 in eye loss during development (Jeffery, 2005; Yamamoto et al., 2004). Alternatively, cryptic variation may

49 be maintained in normal conditions, and expressed as blindness only in case of stress, such as entry into

50 the cave (Rohner et al., 2013). When the cryptic variation is "unmasked" it is then exposed to selection

51 and could become fixed in the population.

52 These mechanisms of selection would result in the evolution of a blind population occurring quite slowly.

53 Furthermore, given that there is often migration from surface to cave populations and that and these

54 populations can interbreed, it seems that blind phenotypes should be lost (Avise and Selander, 1972). One

55 possibility is that the strength of selection for blindness is large enough to counter immigration (Avise and

56 Selander, 1972). Although blind fish maintained in the dark in the lab do not appear to have an advantage

57 of this magnitude (Sadoglu, 1967), recent work suggests a very high cost to developing neural tissue,

58 including eyes (Moran et al., 2015).

59 Due to the immigration of individuals from the surface and the expected level of selection for the eyeless

60 phenotype, cave populations are an example of migration–selection balance (Wright, 1969; Hedrick, 2011;

61 Nagylaki, 1992). However, much of the work in this area has explored the "invasion" of a novel allele or the

62 maintenance of polymorphism, rather than fixation of different alleles in different populations (Yeaman

63 and Otto, 2011; Yeaman and Whitlock, 2011).

64 Here we have developed a model that shows just how strong selection must be to generate blind popula-

65 tions of a cave-dwelling species. Incorporating genetic drift into the model actually increases the level

66 of selection required for fixation. This level of selection is not compatible with the hypothesis that eyes

67 are lost due to drift or that eyes are lost due to selection for improved foraging and pleiotropy (assuming

68 pleiotropy imposes only weak selection). However, if eye development imposes a high cost (Moran et al.,

69 2015) then the adaptation hypothesis is plausible. Alternatively, we suggest reconsidering the historic

70 hypothesis that a high level of selection is due to migration of seeing individuals, who are strongly pho-

71 totactic, out of the cave (Lankester, 1925; Romero, 1985). Thus, we suggest that a standing presence of

72 blindness alleles, combined with extreme loss of sighted individuals in the cave, is a likely scenario leading

73 to evolution of blind cave-dwelling populations.

Model and Analysis

Assumptions

76 Consider a species with two populations: surface-dwelling and cave-dwelling. We are interested in

77 determining when the cave population will evolve blindness, i.e. become comprised of mostly blind

78 individuals, as has occurred in numerous natural systems. We first assume that the surface and cave

79 populations do not experience drift (i.e. populations are of infinite size). Additionally, immigration from

80 the surface population into the cave affects the allele frequency in the cave, but emigration from the

81 cave to the surface does not affect the surface population, as we assume that the surface population is

82 significantly larger than the cave. Generations are discrete and non-overlapping, and mating is random.

83 We track a single biallelic locus, where b^+ is the dominant, seeing allele, and where b^- is the recessive,

84 blindness allele. The frequencies of b^- are denoted by $\tilde{q} \in [0, 1]$ on the surface and $q \in [0, 1]$ in the cave.

85 On the surface, we assume that blindness is selected against, and $\tilde{q}$ is dictated by mutation-selection

86 balance.

Calculating the frequency of the blindness allele

88 Within the cave, the life cycle is as follows. (1) Embryos become juveniles and experience constant selection

89 with relative fitnesses of $w_{b^- b^-} = 1 + s$ and $w_{b^+ b^-} = w_{b^+ b^+} = 1$, where $s \geq 0$. (2) Juveniles migrate into

90 and out of the cave such that a fraction m of adults come from the surface and $1 - m$ from the cave, where

91 $0 \leq m \leq 1$. (3) Adults generate gametes with one-way mutation, where $0 \leq u \leq 1$ is the probability that a

92 b^+ becomes a non-functional b^-. (4) Gametes unite randomly to produce embryos. Given this life cycle,

4

we calculate the allele frequency of the daughter generation (q') via standard equations:

$$q_j = \frac{(1 + s)q^2 + 1q(1 - q)}{(1 + s)q^2 + 1(1 - q^2)} \qquad \text{selection} \qquad (1a)$$

$$q_a = q_j(1 - m) + \tilde{q}m \qquad \text{immigration} \qquad (1b)$$

$$q' = q_a + (1 - q_a)u \qquad \text{mutation} \qquad (1c)$$

The change in allele frequency in one generation is

$$\Delta q = q' - q = \frac{sq^2\left[1 - q - m(1 - u)(1 - \tilde{q})\right] + \left[\tilde{q}m(1 - u) + u - q\left(m(1 - u) + u\right)\right]}{1 + sq^2} \qquad (2)$$

Furthermore, b^- is maintained at a stable equilibrium on the surface: $\tilde{q} = \sqrt{u/z}$, where z is the selection coefficient against b^- on the surface and $u \leq z \leq 1$.

Identifying equilibrium frequencies of the blindness allele

The model we have developed is an example of migration-selection balance (Figure 1; Wright, 1969; Hedrick, 2011; Nagylaki, 1992). An equilibrium exists for this model when $\Delta q = 0$. Assuming $s > 0$ and setting $\Delta q = 0$, Equation 2 can be rearranged into the following cubic polynomial

$$-q^3 + Bq^2 + Cq + D = 0 \qquad (3)$$

where

$$B = 1 - m(1 - u)(1 - \tilde{q})$$
$$C = -\frac{m(1 - u) + u}{s}$$
$$D = \frac{\tilde{q}m(1 - u) + u}{s}$$

5

101 There are three possible roots of this equation, corresponding to three possible equilibria. Depending

102 on the parameter values, Equation 3 may have three real roots or one real root and two imaginary roots.

103 While the values of the roots of this polynomial can be expressed analytically, these equations are too

104 complex to be helpful for understanding the system. For simplicity, we will let $\hat{q}$ represent any possible

105 equilibrium, and $\hat{q}_a \leq \hat{q}_b \leq \hat{q}_c$, stand for the roots of Equation 3.

Rather than tackling the equilibria directly, we first demonstrate that the cave has a protected poly-
morphism. A protected polymorphism exists if the allele frequency moves away from both fixation and
extinction, i.e. $\Delta q < 0$ for $q = 1$ and $\Delta q > 0$ for $q = 0$. For $q = 0$

$$\Delta q = \tilde{q}m(1-u) + u = (1-u)m\sqrt{u/z} + u$$

and $q = 0$ will be an equilibrium if $u = 0$; otherwise $\Delta q > 0$ at $q = 0$ due to immigration of individuals
containing b^- (Figure 2). For $q = 1$

$$\Delta q = -m(1-\tilde{q})(1-u) = -m(1-\sqrt{u/z})(1-u)$$

106 and $q = 1$ will be an equilibrium if $m = 0$ or $u = 1$; otherwise $\Delta q < 0$ at $q = 1$ due to immigration of

107 individuals containing b^+ (Figure 2). Thus a protected polymorphism always exists except at the edge cases

108 $m = 0, u = 0$, and $u = 1$. In biological terms, the cave population will be polymorphic despite directional

109 selection for b^- if there is some immigration from the surface population and the surface population

110 is polymorphic. For $s = 0$, there is only one equilibrium, and it is near 0. For large s, there is only one

111 equilibrium, and it is near 1. Three equilibria will only exist for moderate levels of selection (Figure 2).

112 **Validity of equilibria.**   An equilibrium is only valid in our model if it is real and between $[0, 1]$; otherwise,

113 it is not biologically interpretable in this system. The lower bound for any equilibrium is $\frac{m\tilde{q}(1-u)+u}{m(1-u)+u}$ if

114 $m > 0$ or $u > 0$; otherwise it is 0 (Lemma 1). The upper bound for any equilibrium is $1 - m(1-u)(1-\tilde{q})$

115 (Lemma 2). Thus if any equilibrium is real it is valid. The only exception to this rule is the edge case when

116 $s = m = u = 0$. In this case, all evolutionary forces are eliminated, and $q' = q$ for all q. Here every possible

117 value of q is an equilibrium, although only $q \in [0, 1]$ makes any sense. Furthermore, it is important to

118 note that if $m > 0$,

$$\tilde{q} \leq \frac{m\tilde{q}(1-u)+u}{m(1-u)+u} \implies \tilde{q} \leq \hat{q} \tag{4}$$

119 indicating that the equilibrium frequencies in the cave are always greater than the allele frequency on the

120 surface.

121 **Approximations.** In order to study equilibria, we will simplify our model by assuming that $u \ll 1$ such

122 that $1 - u \approx 1$ and

$$\Delta q \propto sq^2 \left[1 - q - m(1 - \tilde{q})\right] + \left[\tilde{q}m + u - q\left(m + u\right)\right] \tag{5}$$

123 **Weak-selection approximation.** If selection is weak, then an equilibrium exists near $\tilde{q}$ (Figure 2). We

124 use a second-order Taylor series at $q = 0$ to determine the upper bound on s for the presence of three

125 equilibria (i.e. when selection is so strong that an equilibrium near $\tilde{q}$ does not exist). The second-order

126 series allows us to determine the lower two equilibrium points, although this approximation is inaccurate

127 as q increases. This approximation gives us

$$\Delta q \approx s(1-m)q^2 - (m+u)q + m\tilde{q} + u \tag{6}$$

after assuming that $1 - \tilde{q} \approx 1$. This equation has two roots, which are the lowest two of three total equilibria,

$$\hat{q}_{a,1} = \frac{m + u - \sqrt{(m+u)^2 - 4s(1-m)(m\tilde{q}+u)}}{2s(1-m)}$$

$$\hat{q}_{b,1} = \frac{m + u + \sqrt{(m+u)^2 - 4s(1-m)(m\tilde{q}+u)}}{2s(1-m)}$$

128 These two roots exist only if

$$0 < \sqrt{(m+u)^2 - 4s(1-m)(m\tilde{q}+u)} \implies s < \frac{(m+u)^2}{4(1-m)(m\tilde{q}+u)} \tag{7}$$

129 which provides us with an estimate of the upper bound on s for the presence of three equilibria.

130   The derivative of Equation 6 is $\frac{\mathrm{d}\Delta q}{\mathrm{d}q}(q) = 2s(1-m)q - (m+u)$, and a equilibrium will be stable if

131   $-2 < \frac{\mathrm{d}\Delta q}{\mathrm{d}q}(\hat{q}) < 0$. From this, it can be easily shown that $\hat{q}_{a,1}$ is stable and $\hat{q}_{b,1}$ is unstable.

132   **Strong-selection approximation.**   In order to determine the lower bound on s for the presence of three

133   equilibria, we assume that selection is strong enough such that $u/s \approx 0$ and $\tilde{q}/s \approx 0$. Therefore,

$$\Delta q \propto -q\left[q^2 - \left[1 - m(1-\tilde{q})\right]q + m/s\right] \tag{8}$$

and the equilibria can be described as

$$\hat{q}_{a,2} = 0$$

$$\hat{q}_{b,2} = \frac{1}{2}\left(1 - m(1-\tilde{q}) - \sqrt{\left[1 - m(1-\tilde{q})\right]^2 - \frac{4m}{s}}\right)$$

$$\hat{q}_{c,2} = \frac{1}{2}\left(1 - m(1-\tilde{q}) + \sqrt{\left[1 - m(1-\tilde{q})\right]^2 - \frac{4m}{s}}\right)$$

134   The latter two equilibria will exist only if

$$s > \frac{4m}{\left[1 - m(1-\tilde{q})\right]^2}$$

135   which provides us an estimate of the lower bound for the presence of three equilibria.

136   The derivative of Equation 8 is $\frac{\mathrm{d}\Delta q}{\mathrm{d}q}(q) = -3q^2 + 2\left[1 - m(1-\tilde{q})\right]q - m/s$, and it can be easily shown

137   that $\hat{q}_{b,2}$ is unstable and $\hat{q}_{c,2}$ is stable.

138   **Validity of approximations.**   By substituting $\hat{q}_{a,1}$ and $\hat{q}_{b,1}$ back into Equation 5, we obtain $\Delta q = -sq^2(\hat{q} - \tilde{q}m)$.

139   Thus, $\Delta q \le 0$, which indicates that $\hat{q}_{a,1}$ overestimates $\hat{q}_a$ and that $\hat{q}_{b,1}$ underestimates $\hat{q}_b$. By substituting

140   $\hat{q}_{b,2}$ and $\hat{q}_{c,2}$ back into Equation 5, we find that $\Delta q = \tilde{q}m + u(1-q)$. Thus $\Delta q \ge 0$, which indicates that

141   $\hat{q}_{b,2}$ overestimates $\hat{q}_b$ and that $\hat{q}_{c,2}$ underestimates $\hat{q}_c$. However, the error in our approximations is slight

142   (Figure 3).

Dynamics. The dynamics of the evolution of the cave population depend on the parameter values and the starting allele frequency, q_0 (Table 1). If there is one equilibrium value, then b^- will evolve to be the major allele in the population if $\hat{q} > 0.5$. If there are three equilibria, and $\hat{q}_c > 0.5$, then b^- will become the major allele only if its initial frequency is above the threshold $\hat{q}_b$.

Based on these approximations the dynamics of the system can be summarized as follows. First, there are three possible equilibria: $\hat{q}_a \approx \hat{q}_{a,1}, \hat{q}_b \in \left[\hat{q}_{b,1}, \hat{q}_{b,2}\right]$, and $\hat{q}_c \approx \hat{q}_{c,2}$. Second, there are four possible equilibria configurations: 1, 2a, 2b, and 2c.

Case 1, $\frac{(m+u)^2}{4(1-m)(m\tilde{q}+u)} < \frac{4m}{\left[1-m(1-\tilde{q})\right]^2}$: only one equilibrium exists, and it is stable. The population will always evolve towards it.

Case 2, $\frac{4m}{\left[1-m(1-\tilde{q})\right]^2} < \frac{(m+u)^2}{4(1-m)(m\tilde{q}+u)}$: depending on the strength of s, this case may have one of three possible configurations:

Case 2a, $0 \leq s < \frac{4m}{\left[1-m(1-\tilde{q})\right]^2}$: Only one equilibrium exists, $\hat{q}_a$, and it is stable. The population will always evolve towards it.

Case 2b, $\frac{4m}{\left[1-m(1-\tilde{q})\right]^2} < s < \frac{(m+u)^2}{4(1-m)(m\tilde{q}+u)}$: All three equilibria exist; $\hat{q}_a$ and $\hat{q}_c$ are stable, while $\hat{q}_b$ is unstable. If the population starts below $\hat{q}_b$, it will evolve towards $\hat{q}_a$. If it starts above $\hat{q}_b$, it will evolve towards $\hat{q}_c$.

Case 2c, $\frac{(m+u)^2}{4(1-m)(m\tilde{q}+u)} < s$: only one equilibrium, $\hat{q}_c$, exists, and it is stable. The population will always evolve towards it.

The evolution of blindness

When the cave population is founded, its initial allele frequency will likely match the equilibrium frequency on the surface ($q_0 = \tilde{q}$). Since $\tilde{q} < \hat{q}$, the allele frequency in the cave will increase due to selection until it reaches the lowest equilibrium. If this equilibrium is $\geq \frac{1}{2}$, we consider the population to have evolved blindness. If there is only a single equilibrium, and $s \geq \frac{4m-8\tilde{q}m-4u}{1-2m(1-\tilde{q})}$ (i.e. $m \leq \frac{s+4u}{2s(1-\tilde{q})+4-8\tilde{q}}$, Lemma 3), b^- will evolve to become the major allele. If there are three equilibria (Case 2b), then the population will not

9

167 evolve blindness: the maximum value of $\hat{q}_a \approx \frac{1}{8}$. Therefore, blindness will only evolve if

$$s \geq \max \left\{ \frac{4m - 8\tilde{q}m - 4u}{1 - 2m(1 - \tilde{q})}, \frac{(m + u)^2}{4(1 - m)(m\tilde{q} + u)} \right\} \approx s \geq \frac{m}{4\left(\tilde{q} + \frac{u}{m}\right)} \tag{9}$$

168 This approximation is valid when $\tilde{q} < \frac{1}{16}$ and $u \ll m \ll 1$. We analytically calculated ultimate allele

169 frequencies for our model and compared it to the above approximation (Figure 4A), and we also explored

170 the approximation when u and $\tilde{q}$ are varied (Figure 5).

Finite-population simulations

Constant migration

173 To investigate the impact of drift on our model, we simulated diploid populations of size $N = 1000$, where

174 the frequency of adults was determined by drawing $2N$ alleles from a binomial distribution with mean q_a

175 (Equation 1b). q' was calculated based on the post-drift adult allele frequency and immigration. For each

176 set of parameters, we recorded the average q' frequency across these 100 populations at specific time

177 points.

178 For high migration rates, the average allele frequency is similar to the infinite model, except that drift allows

179 some populations that have three equilibria to evolve blindness (Figure 4B). However, at low migration

180 rates ($m < u/\tilde{q} = 10^{-4}$), populations have low average frequency of b^- at 10 thousand generations,

181 unless $s > 1$. As immigration decreased, these populations became dependent on *de novo* mutations

182 to produce b^-, which is a slow process. At 5 million generations, which is close to the estimated age of

183 cavefish populations Gross (2012), the average allele frequency is a better match to the results from the

184 the infinite-population model (Figure 4C); although, selection is ineffective for $s < 1/2N = 5 \times 10^{-4}$.

Episodic migration

186 Because cave and surface populations are may be connected intermittently due to flooding, we simulated

187 periods of immigration followed by periods of isolation following a first-order Markov process. The

188 probability of switching between isolation and immigration or vice versa was 10% every generation.

189 Results for the intermittently connected simulations were nearly identical to previous simulations, with

190 the exception that at high levels of migration and selection, drift was more effective in increasing allele

191 frequencies (Figure 4D).

Discussion

193 Both our model and simulations show that strong selection (characterized as $s > 0.05$; Rieseberg and

194 Burke, 2001) is necessary for a cave population to evolve blindness. Our simulations demonstrate that

195 genetic drift, which is likely to occur in small cave populations, markedly decreases the fixation of these

196 rare alleles, resulting in the need for even greater selection.

Model

198 Our model demonstrates that without drift, blindness occurs in cave populations only when s is large or

199 m is very small. This result is logical: given low levels of immigration, selection increases the frequency

200 of the blindness allele. In contrast, for high levels of immigration there will always be some sighted fish.

201 This case essentially results in a single population, for which selection in the large surface population for

202 sighted fish outweighs small-scale selection in the cave (Nagylaki and Lou, 2008). In other words, gene

203 flow prevents local adaptation, as expected. For intermediate levels of immigration, there will be sighted

204 fish unless selection removes them.

205 What is surprising about our result is the level of selection required to fix blindness in the population. This

206 result contrasts with the level of the selection found in most cases of local adaptation. Sadoglu (1967)

207 argued that given the observed number of populations of blind fish living in caves, drift would have fixed

208 at least one for a useful eye, However, because all populations appear to be blind or with a significantly

209 reduced eye, this indicates strong selection for "degenerative genes".

Drift

The "neutral mutation hypothesis" is equivalent to no selection for the blindness allele in a small population. This hypothesis relies on drift to fix populations for blindness. Thus, we explored the interaction of immigration, selection and blindness with simulations. However, our simulations including drift produce similar results to our model, with two notable exceptions. First, as observed previously, drift removes the blindness allele when it occurs at low levels; thus, for low immigration rates, populations consisted primarily of sighted fish, except when very strong selection immediately increased the frequency of blindness. This is opposite of predicted by the "neutral mutation hypothesis".

When migration is very high, drift has minimal effect, as the two populations essentially function as one, with the surface population swamping the smaller cave population. However, for intermediate levels of migration the level of selection required to fix blindness decreased substantially. In this case, immigration is increasing the frequency of the blindness allele in each generation, allowing more chances for selection to overcome drift. This result is consistent with the observation of Blanquart et al. (2012)

Allowing populations to evolve for longer periods of time does increase the likelihood that a population can drift to fixation for blindness. As for higher levels of immigration, more generations results in a greater chance for a blindness mutation to occur in the population, and for drift to increase the frequency of this allele. Similarly, increasing the mutation rate, either for the whole genome or as a way to allow multiple mutations to produce blindness, results in a greater likelihood of blind populations. As for higher levels of immigration, a greater chance of producing a mutation for blindness results in an increased number of chances for a population to evolve blindness.

Effect of intermittent connections

On the low end of migration, intermittent connections effectively result in a decrease in the immigration rate and "replenishment" of the blindness allele, which increases in the level of selection required to fix blindness in the population. In contrast, when immigration rates are high, disconnecting the two populations and effectively reducing the immigration rate allows populations to fix for blindness, at least until the next period of high immigration.

Previous observations of strong selection

The values we suggest here as "strong selection" are high, but not inconsistent with previous observations. Previous calculations of strong selection resulting in selective sweeps in wild populations range from 0.02–0.7 (Sáez et al., 2003; Schlenke and Begun, 2004; Wootton et al., 2002; Nair et al., 2003). Estimated selection coefficients for drug resistance in *Plasmodium falciparum* were 0.1–0.7 (Wootton et al., 2002; Nair et al., 2003). These extremely high values for selection led to fixation in 20-80 generations. For a major advantageous allele, the average value of s has been estimated as 0.11 in plants and 0.13 in animals (Rieseberg and Burke, 2001; Morjan and Rieseberg, 2004). Thus, estimated selection coefficients for cave species are consistent with a selection mechanism that is stronger than previously proposed.

Potential mechanisms of strong selection

Previous mechanisms of selection proposed for cave species have primarily been weak (Darwin, 1859; Sadoglu, 1967). However, recent work has suggested that eye development imposes a high metabolic cost, particularly for juveniles (Moran et al., 2015). In a food-limited environment, like a cave, this cost could lead to the level of selection suggested by our model. Additionally, an alternative mechanism of selection exists: migration of seeing individuals, who are strongly phototactic, out of the cave (Lankester, 1925; Romero, 1985). Emigration of sighted individuals functions like selection because it systematically removes b^+ alleles from the cave population. Phototaxis has been observed in eyed cavefish (Espinasa and Borowsky, 2000). Migration thus imposes strong selection in the cave for blind individuals. This mechanism of strong selection would explain the observation that blindness in cave dwelling organisms evolved repeatedly. This hypothesis is consistent with all previous hypotheses of how blindness arises (i.e. random mutation or differential expression) but suggests an alternative mechanism of selection that is much stronger than previously proposed fitness advantages (i.e. reallocation of resources). Furthermore, it suggests a way of maintaining a mostly blind cave population despite interbreeding and gene flow from surface populations (Bradic et al., 2012; Avise and Selander, 1972). Interestingly, our work is consistent with work suggesting standing cryptic variation for eye size in cavefish (Rohner et al., 2013; Rohner, 2015). The primary effect of the variation being cryptic rather than *de novo* would be that the allele frequency in the surface population would be higher than expected. However, strong selection would still be required.

13

263 Alternatively, vibration attraction behavior, where individuals are attracted to moving objects, provides a

264 strong advantage for some individuals by allowing them to find food (Yoshizawa et al., 2012). This behavior

265 is observed in multiple cavefish populations (Yoshizawa et al., 2010). In contrast, this behavior would like

266 result in predation in a sighted environment (Yoshizawa and Jeffery, 2011). Surface populations show a

267 low frequency of this behavior (Yoshizawa and Jeffery, 2011). However, in this case the genetic basis of the

268 behavior would have to be linked to blindness, and whether this is the case is unknown.

Application to other local adaptation scenarios

270 Although we have described our model in the context of cave and surface populations, with the allele

271 under selection for blindness, this work also applies to other scenarios of local adaptation. Generally, our

272 scenario can be considered a metapopulation with divergent selection (Blanquart et al., 2012; Yeaman

273 and Otto, 2011). However, while previous work focused on stable polymorphism, here we have addressed

274 how populations become fixed for a state.

Acknowledgments

Appendix

279 **Lemma 1.** *If $m > 0$ or $u > 0$, and $s \geq 0$, the minimum value of an equilibrium is $\frac{m\tilde{q}(1-u)+u}{m(1-u)+u}$. If $m = u = 0$*

280 *and $s > 0$, the minimum value of an equilibrium is 0. If $m = u = s = 0$, all points are equilibria, and thus 0*

281 *is the lowest possible valid equilibrium.*

282 *Proof.* Case 1. Let $f(q)$ represent the change in allele frequency over one generation (Equation 2). Let

283 $q^* = \frac{m\tilde{q}(1-u)+u}{m(1-u)+u}$. If $s = 0$ and $m > 0$ (or $u > 0$), $f(q^*) = 0$, and therefore q^* is an equilibrium for these

14

284 parameters.

285 Now let $s \geq 0$, and $q < q^*$. The denominator of $f(q)$ is positive, and the numerator is

$$sq^2\left[1 - q - m(1-u)(1-\tilde{q})\right] + \left[\tilde{q}m(1-u) + u - q\left(m(1-u)+u\right)\right] \tag{10}$$

$$q < \frac{m\tilde{q}(1-u)+u}{m(1-u)+u} \implies q < 1 - m(1-u)(1-\tilde{q})$$

$$\iff 1 - q - m(1-u)(1-\tilde{q}) > 0$$

$$\iff \text{the first part of (10) is positive}$$

$$q < \frac{m\tilde{q}(1-u)+u}{m(1-u)+u} \iff q\left(m(1-u)+u\right) < \tilde{q}m(1-u)+u$$

$$\iff \tilde{q}m(1-u) + u - q\left(m(1-u)+u\right) > 0$$

$$\iff \text{the second part of (10) is positive}$$

$$\therefore \text{(10) is positive, and } f(q) > 0 \; \forall q < q^*$$

286 Since $f(q) > 0 \; \forall q < q^*$, q^* is the lowest value that can be an equilibrium.

287 Case 2. Let $m + u = 0$ and $s > 0$. Now (10) is $sq^2(1-q)$, which has its lowest equilibrium at 0.

288 Case 3. If $s = 0$, all possible values of q are equilibria. □

289 **Lemma 2.** *The maximum value of an equilibrium is* $1 - m(1-u)(1-\tilde{q})$.

290 *Proof.* Let $q^* = 1 - m(1-u)(1-\tilde{q})$. Since $\lim_{s\to\infty} f(q^*) = 0$, q^* is a potential equilibrium. Now let

291 $q > q*$.

15

$$q > 1 - m(1-u)(1-\tilde{q}) \iff 1 - q - m(1-u)(1-\tilde{q}) < 0$$

$$\iff \text{the first part of (10) is negative}$$

$$q > 1 - m(1-u)(1-\tilde{q}) \iff q > 1 - m(1-u) + \tilde{q}m(1-u)$$

$$\iff q - 1 + m(1-u) + u > \tilde{q}m(1-u) + u$$

$$\iff q(m(1-u) + u) - (1-q)(1 - m(1-u) - u) > \tilde{q}m(1-u) + u$$

$$\implies q(m(1-u) + u) > \tilde{q}m(1-u) + u$$

$$\iff \text{the second part of (10) is negative}$$

$$\therefore \text{(10) is negative, and } f(q) < 0 \; \forall q > q^*$$

Since $f(q) < 0 \; \forall q > q^*$, q^* is the highest value that can be an equilibrium. $\qquad\square$

Lemma 3. *If* $m \leq \frac{s+4u}{(2s(1-\tilde{q})+4-8\tilde{q})(1-u)}$ *and* $\tilde{q} < 1/2$, *there exists an equilibrium* $\geq 1/2$.

Proof. By rearranging Equation 3, we find a formula for the migration rate that will generate an equilibrium at q:

$$\hat{m}(q) = \frac{(1-q)(q^2s + u)}{(q^2s(1-\tilde{q}) + q - \tilde{q})(1-u)} \tag{11}$$

First, we can show that $\hat{m}(q)$ monotonically decreases for $q \geq 1/2$:

$$-\frac{\mathrm{d}\hat{m}(q)}{\mathrm{d}q} = sq\left(2\tilde{q} + 2(1-\tilde{q})u - \left(1 + 3\tilde{q} + (1-\tilde{q})u\right)q + 2q^2 + s(1-\tilde{q})q^3\right) + (1-\tilde{q})u$$

And

$$0 \leq \tilde{q} \leq 1/2 \leq q \leq 1 \implies (2q-1)(q-\tilde{q}) + \tilde{q}(1-q) + (2-q)(1-\tilde{q})u \geq 0$$

$$\iff 2\tilde{q} + 2(1-\tilde{q})u - \left(1 + 3\tilde{q} + (1-\tilde{q})u\right)q + 2q^2 \geq 0$$

$$\implies sq\left(2\tilde{q} + 2(1-\tilde{q})u - \left(1 + 3\tilde{q} + (1-\tilde{q})u\right)q + 2q^2 + s(1-\tilde{q})q^3\right)$$

$$+ (1-\tilde{q})u \geq 0$$

$$\iff \frac{\mathrm{d}\hat{m}(q)}{\mathrm{d}q} \leq 0$$

16

Therefore, $\hat{m}(q)$ monotonically decreases from $\hat{m}(1/2)$ to $\hat{m}(1) = 0$ as q increases from $1/2$ to 1. Stated another way, if $0 \leq m \leq \hat{m}(1/2)$ there exists an equilibrium $1/2 \leq q \leq 1$. $\qquad\square$

References

Avise, J. C. and R. K. Selander, 1972. Evolutionary genetics of cave-dwelling fishes of the genus *Astyanax*. Evolution 26:1–19.

Barr, T. C., 1968. Cave ecology and the evolution of troglobites. chap. 2, Pp. 35–102, *in* T. Dobzhansky, M. Hecht, and W. Steere, eds. Evolutionary Biology, vol. 2. Springer, New York.

Blanquart, F., S. Gandon, and S. L. Nuismer, 2012. The effects of migration and drift on local adaptation to a heterogeneous environment. J. Evol. Biol. 25:1351–1363.

Borowsky, R., 2015. Regressive evolution: Testing hypotheses of selection and drift. chap. 5, Pp. 93–109, *in* A. Keene, M. Yoshizawa, and S. McGaugh, eds. Biology and Evolution of the Mexican Cavefish. Academic Press, San Diego, CA.

Bradic, M., P. Beerli, F. J. García-de León, S. Esquivel-Bobadilla, and R. L. Borowsky, 2012. Gene flow and population structure in the mexican blind cavefish complex (*Astyanax mexicanus*). BMC Evol. Biol. 12:9.

Coghill, L. M., C. D. Hulsey, J. Chaves-Campos, F. J. G. de Leon, and S. G. Johnson, 2014. Next generation phylogeography of cave and surface *Astyanax mexicanus*. Mol. Phylogenet. Evol. 79:368–374.

Culver, D. C., L. L. Master, M. C. Christman, and H. H. Hobbs, 2000. Obligate cave fauna of the 48 contiguous United States. Conserv. Biol. 14:386–401.

Darwin, C., 1859. On the origin of species based on natural selection, or the preservation of favoured races in the struggle for life. 6th ed. John Murray, London.

Dowling, T. E., D. P. Martasian, and W. R. Jeffery, 2002. Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. Mol. Biol. Evol. 19:446–455.

Espinasa, L. and R. Borowsky, 2000. Eyed cave fish in a karst window. J. Cave Karst Stud. 62:180–183.

320 Gross, J. B., 2012. The complex origin of *Astyanax* cavefish. BMC Evol. Biol. 12:105.

321 Gross, J. B., R. Borowsky, and C. J. Tabin, 2009. A novel role for Mc1r in the parallel evolution of depigmen-
322 tation in independent populations of the cavefish *Astyanax mexicanus*. PLoS Genet. 5:e1000326.

323 Hedrick, P. W., 2011. Gene flow and selection. chap. 7.IV, Pp. 424–436, *in* Genetics of Populations, 4th ed.
324 Jones and Bartlett, Sudbury, MA.

325 Hinaux, H., J. Poulain, C. Da Silva, C. Noirot, W. R. Jeffery, D. Casane, and S. Rétaux, 2013. De novo
326 sequencing of *Astyanax mexicanus* surface fish and Pachón cavefish transcriptomes reveals enrichment
327 of mutations in cavefish putative eye genes. PloS One 8:e53553.

328 Jeffery, W. R., 2001. Cavefish as a model system in evolutionary developmental biology. Dev. Biol. 231:1–12.

329 ———, 2005. Adaptive evolution of eye degeneration in the Mexican blind cavefish. J. Hered. 96:185–196.

330 Jeffery, W. R., A. G. Strickler, and Y. Yamamoto, 2003. To see or not to see: evolution of eye degeneration in
331 mexican blind cavefish. Integr. Comp. Biol. 43:531–541.

332 Kimura, M. and J. L. King, 1979. Fixation of a deleterious allele at one of two" duplicate" loci by mutation
333 pressure and random drift. Proceedings of the National Academy of Sciences 76:2858–2861.

334 Langecker, T. G., H. Schmale, and H. Wilkens, 1993. Transcription of the opsin gene in degenerate eyes of
335 cave-dwelling *Astyanax fasciatus* (Teleostei, Characidae) and of its conspecific epigean ancestor during
336 early ontogeny. Cell Tissue Res. 273:183–192.

337 Lankester, E., 1925. The blindness of cave-animals. Nature 116:745–746.

338 Moran, D., R. Softley, and E. J. Warrant, 2015. The energetic cost of vision and the evolution of eyeless
339 Mexican cavefish. Science Advances 1:e1500363.

340 Morjan, C. L. and L. H. Rieseberg, 2004. How species evolve collectively: implications of gene flow and
341 selection for the spread of advantageous alleles. Mol. Ecol. 13:1341–1356.

342 Nagylaki, T., 1992. Migration and selection. chap. 6, Pp. 128–152, *in* Introduction to Theoretical Population
343 Genetics. Springer-Verlag, Berlin.

Nagylaki, T. and Y. Lou, 2008. The dynamics of migration–selection models. chap. 4, Pp. 117–170, *in* A. Friedman, ed. Tutorials in Mathematical Biosciences IV. Springer, New York.

Nair, S., J. T. Williams, A. Brockman, L. Paiphun, M. Mayxay, P. N. Newton, J.-P. Guthmann, F. M. Smithuis, T. T. Hien, N. J. White, et al., 2003. A selective sweep driven by pyrimethamine treatment in southeast Asian malaria parasites. Mol. Biol. Evol. 20:1526–1536.

Niven, J. E., 2008. Evolution: convergent eye losses in fishy circumstances. Curr. Biol. 18:R27–R29.

Niven, J. E. and S. B. Laughlin, 2008. Energy limitation as a selective pressure on the evolution of sensory systems. J. Exp. Biol. 211:1792–1804.

Protas, M., M. Conrad, J. B. Gross, C. Tabin, and R. Borowsky, 2007. Regressive evolution in the Mexican cave tetra, *Astyanax mexicanus*. Curr. Biol. 17:452–454.

Protas, M. E., C. Hersey, D. Kochanek, Y. Zhou, H. Wilkens, W. R. Jeffery, L. I. Zon, R. Borowsky, and C. J. Tabin, 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat. Genet. 38:107–111.

Rieseberg, L. H. and J. M. Burke, 2001. The biological reality of species: gene flow, selection, and collective evolution. Taxon 50:47–67.

Rohner, N., 2015. Selection through standing genetic variation. *in* A. Keene, M. Yoshizawa, and S. McGaugh, eds. Biology and Evolution of the Mexican Cavefish, Pp. 137–154. Academic Press, San Diego, CA.

Rohner, N., D. F. Jarosz, J. E. Kowalko, M. Yoshizawa, W. R. Jeffery, R. L. Borowsky, S. Lindquist, and C. J. Tabin, 2013. Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. Science 342:1372–1375.

Romero, A., 1985. Ontogenetic change in phototactic responses of surface and cave populations of *Astyanax fasciatus* (Pisces: Characidae). Copeia 1985:1004–1011.

Sadoglu, P., 1967. The selective value of eye and pigment loss in Mexican cave fish. Evolution 21:541–549.

Sáez, A. G., A. Tatarenkov, E. Barrio, N. H. Becerra, and F. J. Ayala, 2003. Patterns of DNA sequence

368 polymorphism at Sod vicinities in *Drosophila melanogaster*: unraveling the footprint of a recent selective
369 sweep. Proc. Natl. Acad. Sci. U. S. A. 100:1793–1798.

370 Schlenke, T. A. and D. J. Begun, 2004. Strong selective sweep associated with a transposon insertion in
371 *Drosophila simulans*. Proc. Natl. Acad. Sci. U. S. A. 101:1626–1631.

372 Strickler, A. G., Y. Yamamoto, and W. R. Jeffery, 2007. The lens controls cell survival in the retina: evidence
373 from the blind cavefish *Astyanax*. Dev. Biol. 311:512–523.

374 Wilkens, H., 1988. Evolution and genetics of epigean and cave *Astyanax fasciatus* (Characidae, Pisces). Pp.
375 271–367, *in* T. Dobzhansky, M. Hecht, and W. Steere, eds. Evolutionary biology. Springer.

376 Wootton, J. C., X. Feng, M. T. Ferdig, R. A. Cooper, J. Mu, D. I. Baruch, A. J. Magill, and X.-Z. Su, 2002. Genetic
377 diversity and chloroquine selective sweeps in *Plasmodium falciparum*. Nature 418:320–323.

378 Wright, S., 1969. Selection and immigration. Pp. 36–38, *in* Evolution and the Genetics of Populations,
379 vol. 2. The University of Chicago Press, Chicago.

380 Yamamoto, Y., D. W. Stock, and W. R. Jeffery, 2004. Hedgehog signalling controls eye degeneration in blind
381 cavefish. Nature 431:844–847.

382 Yeaman, S. and S. P. Otto, 2011. Establishment and maintenance of adaptive genetic divergence under
383 migration, selection, and drift. Evolution 65:2123–2129.

384 Yeaman, S. and M. C. Whitlock, 2011. The genetic architecture of adaptation under migration-selection
385 balance. Evolution 65:1897–1911.

386 Yoshizawa, M., Š. Gorički, D. Soares, and W. R. Jeffery, 2010. Evolution of a behavioral shift mediated by
387 superficial neuromasts helps cavefish find food in darkness. Curr. Biol. 20:1631–1636.

388 Yoshizawa, M. and W. R. Jeffery, 2011. Evolutionary tuning of an adaptive behavior requires enhancement
389 of the neuromast sensory system. Commun. Integr. Biol. 4:89–91.

390 Yoshizawa, M., Y. Yamamoto, K. E. O'Quin, and W. R. Jeffery, 2012. Evolution of an adaptive behavior and
391 its sensory receptors promotes eye regression in blind cavefish. BMC Biol. 10:108.
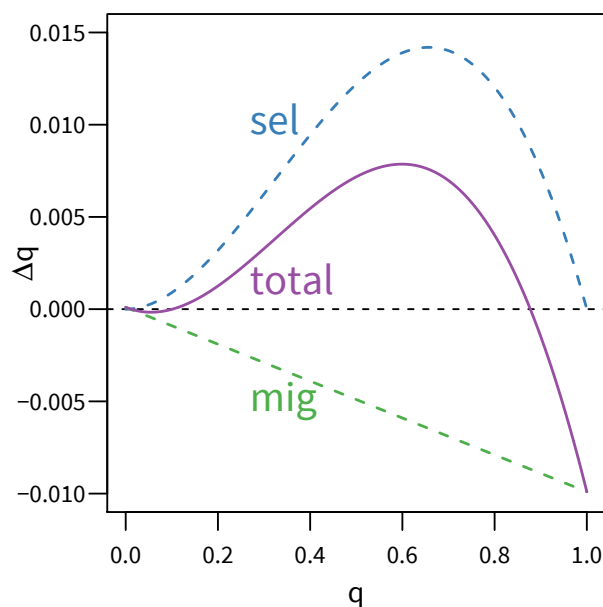
Figures

Figure 1: Dynamics of cave populations are governed by migration-selection balance. The top curve shows the change in allele frequency due to selection, while the bottom curve shows the change in allele frequency due to immigration. The middle curve shows how selection and migration are balanced to produced three equilibria points. Parameters: $s = 0.1, m = 0.01, u = 10^{-6}, \tilde{q} = 0.01$.
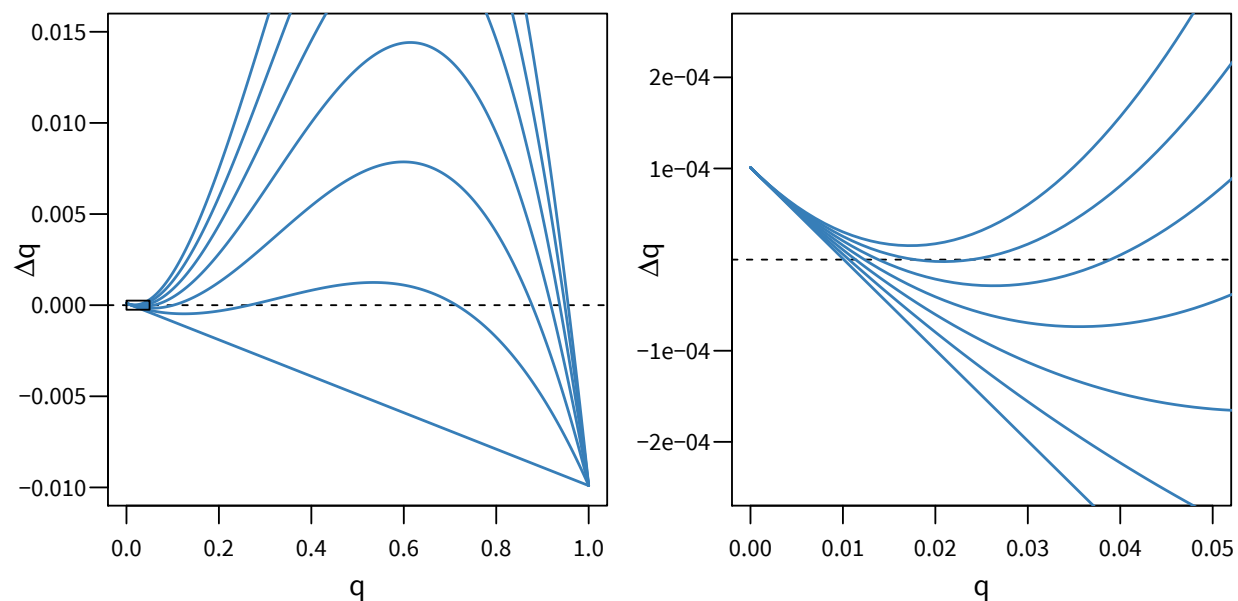
Figure 2: As selection increases, the evolutionary dynamics of the cave population changes. When s is low (bottom line; $s = 0$), there is only one equilibrium: near 0. As s increases (middle five lines, $s = 0.05, 0.1, 0.15, 0.2,$ and 0.25) the local maximum (upper hump) increases and crosses the x-axis, producing three equilibria. When s gets high enough (top line; $s = 0.3$), the local minimum (lower valley) also crosses the x-axis, resulting in one equilibrium again. For all curves $m = 0.01, u = 10^{-6},$ and $\tilde{q} = 0.01$. The figure on the right is an enlarged view of the box in the figure on the left.
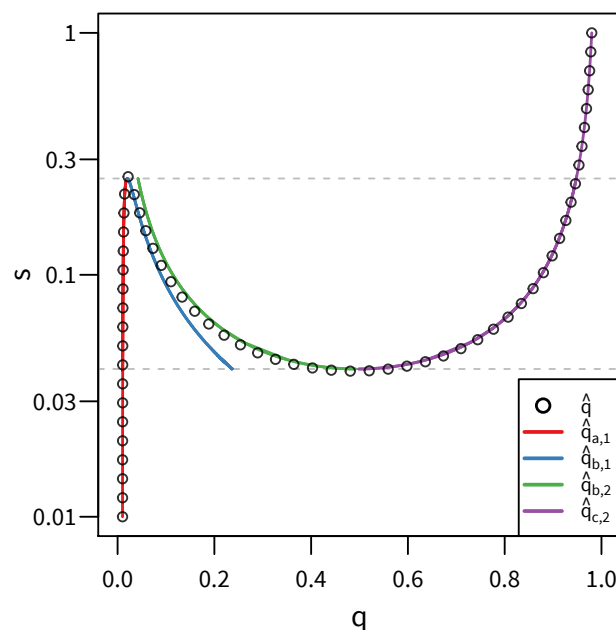
Figure 3: Our equilibria approximations are accurate. The approximations developed in this paper (solid lines) are a good fit for calculated values of s that result in equilibrium for a given q (circles) using Equation 3. Because these approximations are accurate we can use them to better understand this system. The dashed lines are our approximate bounds for the existence of three equilibria (i.e. for small and large values of s, there is one value of q; for intermediate values of s there are three possible values of q). Other parameters are $m = 0.01$, $u = 10^{-6}$, and $\tilde{q} = 0.01$.
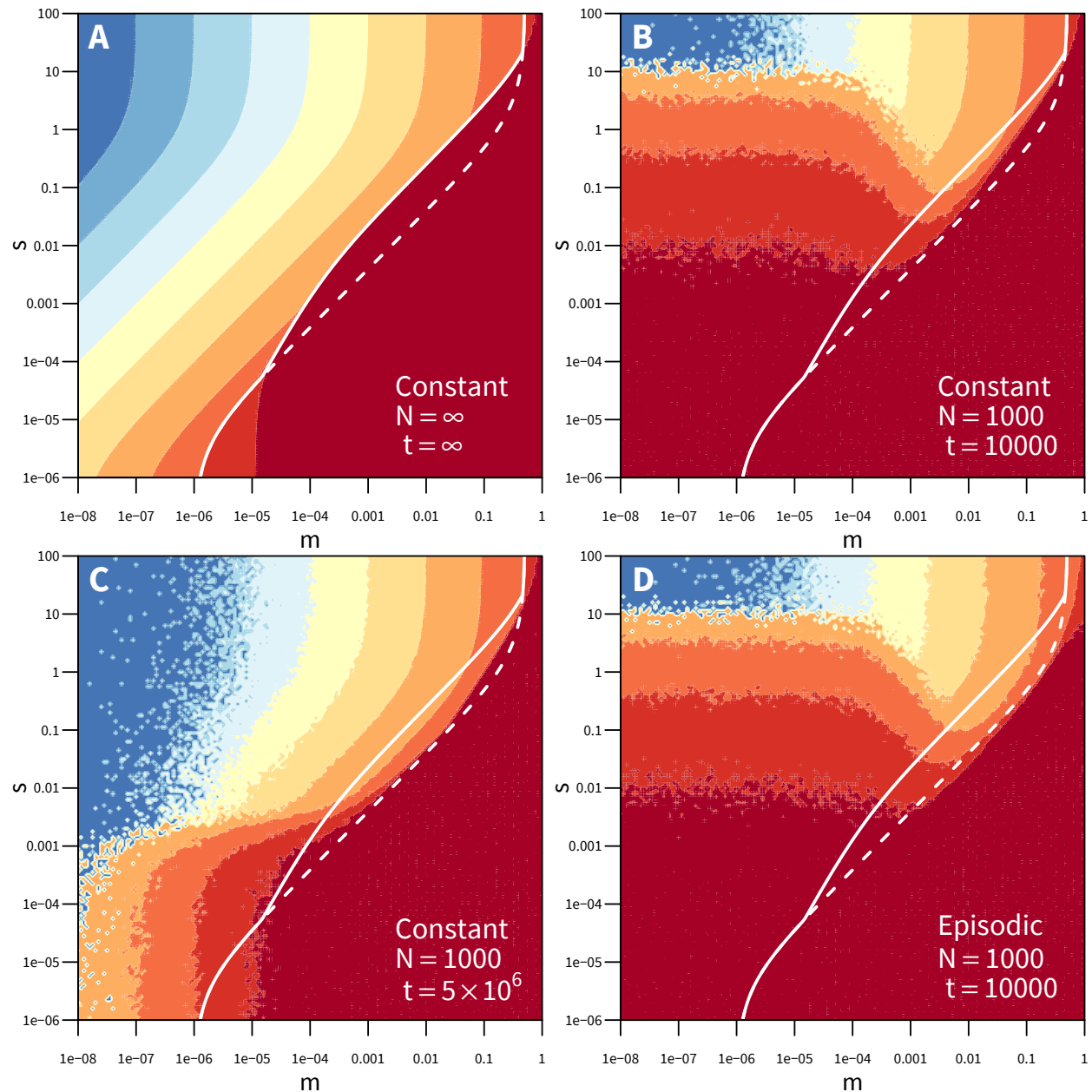
Figure 4: Populations evolve blindness with the help of strong selection. (A) The equilibrium b^- frequency for an infinite population and (B–D) average frequencies after t generations in finite populations (100 replicates). Colors correspond to the frequency of b^- for a given combination of s and m, where blue is high (blindness evolved) and red is low (blindness did not evolve). Colors levels are determined by a base-10 logit scale. The solid white line corresponds to the threshold calculated in Equation 9, and the dash white line corresponds to $\hat{m}(1/2)$ (Equation 11). The area between the solid and dashed lines corresponds to the region where three equilibria exist. Other parameters are $u = 10^{-6}$, $\tilde{q} = 0.01$, and $q_0 = \tilde{q}$.
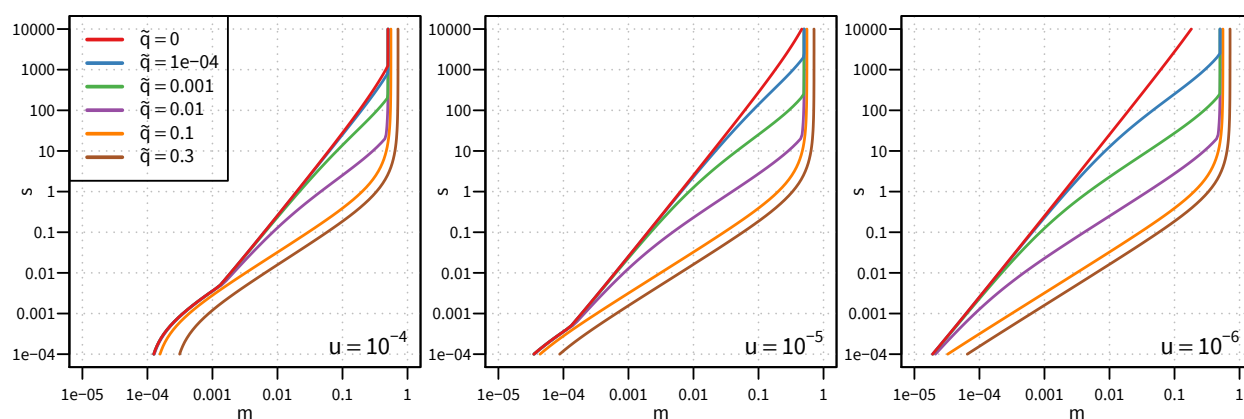
Figure 5: Strong selection is necessary for the evolution of blindness. Each panel represents a different mutation rate, and each line represents a threshold for the evolution of blindness based on $\tilde{q}$ and $q_0 = \tilde{q}$. The rate of mutation of b^+ to b^- alleles is a significant factor in the establishment of the thresholds, e.g. if $\tilde{q} = 0.01$ and $m = 0.01$, then $s > 0.129$ for $u = 10^{-4}$ and $s > 0.250$ for $u = 10^{-6}$; if $m = 0.001$, then $s > 0.004$ for $u = 10^{-4}$ and $s > 0.023$ for $u = 10^{-6}$.

Tables

Table 1: Evolutionary Dynamics of Blindness as $t \to \infty$

One Equilibrium

$0 \le q_0 < \hat{q}$ q_t increases to $\hat{q}$

$\hat{q} < q_0 \le 1$ q_t decreases to $\hat{q}$

Three Equilibria

$0 \le q_0 < \hat{q}_a$ q_t increases to $\hat{q}_a$

$\hat{q}_a < q_0 < \hat{q}_b$ q_t decreases to $\hat{q}_a$

$\hat{q}_b < q_0 < \hat{q}_c$ q_t increases to $\hat{q}_c$

$\hat{q}_c < q_0 \le 1$ q_t decreases to $\hat{q}_c$