

Strong Selection is Necessary for Evolution of Blindness in Cave Dwellers

Reed A. Cartwright,^{1,2*} Rachel S. Schwartz,¹ Alexandra L. Merry,³ and Megan M. Howell³

¹The Biodesign Institute

²School of Life Sciences

³Barrett, The Honors College

Arizona State University

Tempe, AZ, USA

*Corresponding author: cartwright@asu.edu

Running title: Blindness Evolves in Caves Due to Strong Selection

Key words: migration-selection balance, population genetics, models/simulations, mutations

1 **Abstract**

2 Blindness has evolved repeatedly in cave-dwelling organisms, and investigating the loss of sight in cave
3 dwellers presents an opportunity to understand the operation of fundamental evolutionary processes,
4 including drift, selection, mutation, and migration. The observation of blind organisms has prompted
5 many hypotheses for their blindness, including both accumulation of neutral, loss-of-function mutations
6 and adaptation to darkness. Here we model the evolution of blindness in caves. This model captures the
7 interaction of three forces: (1) selection favoring alleles causing blindness, (2) immigration of sightedness
8 alleles from a surface population, and (3) loss-of-function mutations creating blindness alleles. We
9 investigated the dynamics of this model and determined selection-strength thresholds that result in
10 blindness evolving in caves despite immigration of sightedness alleles from the surface. Our results
11 indicate that strong selection is required for the evolution of blindness in cave-dwelling organisms, which
12 is consistent with recent work suggesting a high metabolic cost of eye development.

13 Introduction

14 Blindness has evolved repeatedly across taxa in caves, creating nearly a thousand cave-dwelling species
15 and many more sub-populations (Culver et al., 2000; Dowling et al., 2002; Bradic et al., 2012; Coghill
16 et al., 2014). Surprisingly, many populations of blind individuals experience some level of immigration,
17 which would be expected to prevent the fixation of blindness in a newly established population (Avisé and
18 Selander, 1972; Bradic et al., 2012; Coghill et al., 2014). Thus, blind cave-dwelling populations of typically
19 sighted species pose an interesting challenge to our understanding of evolutionary biology. Namely, how
20 does significant population differentiation evolve despite homogenizing immigration?

21 Several hypotheses have been put forward to explain the evolution of blindness in cave-dwelling species.
22 Darwin suggested that eyes would be lost by “disuse” (Darwin, 1859). We now consider this hypothesis the
23 “neutral-mutation hypothesis” — random mutations can accumulate in genes or regulatory regions related
24 to sight when, as in caves, there is no purifying selection to eliminate them. However, the accumulation
25 of mutations causing blindness due to mutation pressure would take a long time to result in fixation of
26 blindness in populations on its own (Barr, 1968). Thus, genetic drift has been proposed to accelerate
27 the evolution of blindness due to mutation pressure (Kimura and King, 1979; Borowsky, 2015; Wilkens,
28 1988). This hypothesis of relaxed selection appears to be supported by the observation of a high number
29 of substitutions in putative eye genes in the blind forms of cavefishes (Hinaux et al., 2013; Protas et al.,
30 2006; Gross et al., 2009). However, repeatedly developing blindness in cave populations simply by drift in
31 isolation seems unlikely.

32 Relaxing selection that maintains the eye, however, also allows for other agents of selection to act on
33 this trait (Lahti et al., 2009). The “adaptation hypothesis” suggests that there is a cost to an eye; thus,
34 individuals without eyes have greater fitness when eyes are not helpful, resulting in the eventual
35 elimination of seeing individuals. This cost may either come from the energy required to develop a complex
36 structure or due to the vulnerability of the eye (Barr, 1968; Strickler et al., 2007; Jeffery, 2005; Protas
37 et al., 2007; Niven, 2008; Niven and Laughlin, 2008; Moran et al., 2015). Alternatively, positive selection
38 may act on genes related to the eye if these genes act pleiotropically on traits that are beneficial in the
39 dark. For example, in the Mexican tetra (*Astyanax mexicanus*) increased expression of Hedgehog (Hh)
40 affects feeding structures, allowing better foraging in low light conditions (Jeffery, 2001, 2005). Increased

41 Hh signaling also inhibits pax6 expression, which results in eye loss during development (Yamamoto
42 et al., 2004; Jeffery, 2005). Alternatively, cryptic variation may be maintained in normal conditions and
43 expressed as blindness only in case of stress, such as entry into the cave (Rohner et al., 2013). When the
44 cryptic variation is “unmasked”, it is then exposed to selection and could become fixed in the population.

45 Given that there is often gene flow from surface populations into caves, it seems that blind phenotypes
46 should be lost unless selection for blindness is large (Avisé and Selander, 1972). Recent work suggests a
47 very high cost to developing neural tissue, including eyes (Moran et al., 2015). This cost, combined with
48 pleiotropic effects, could lead to blindness despite immigration. However, the level of selection required
49 to induce blindness in cave populations has not been quantified.

50 Here, we model the effects of migration, selection, and mutation to determine the conditions required for
51 the evolution of blindness. This model allows us to explore migration-selection-mutation balance. Where
52 previous theory have explored this balance more generally (Haldane, 1930; Wright, 1931, 1969; Hedrick,
53 2011; Nagylaki, 1992; Yeaman and Otto, 2011; Yeaman and Whitlock, 2011; Bulmer, 1972), we address cavefish
54 evolution specifically. The amount of selection required to oppose the force of immigration is high, but
55 consistent with previous work on metabolic costs in novel environments and selection in other species.
56 Interestingly, drift only impacts blindness in the cave population in a limited range of combinations of
57 selection, dominance, and migration.

58 **Model and Analysis**

59 **Assumptions**

60 We consider two populations: surface-dwelling and cave-dwelling. We are interested in determining
61 when the cave population will evolve blindness, i.e. become mostly comprised of blind individuals, as
62 has occurred in numerous natural systems. We first assume that the surface and cave populations do
63 not experience drift (i.e. populations are of infinite size). Additionally, immigration from the surface
64 population into the cave affects the allele frequency in the cave, but emigration from the cave to the
65 surface does not affect the surface population, as we assume that the surface population is significantly

66 larger than the cave. Generations are discrete and non-overlapping, and mating is random. We track a
 67 single biallelic locus, where B is the seeing allele and where b is blindness allele. The frequency of b is
 68 denoted by $Q \in [0, 1]$ on the surface and $q \in [0, 1]$ in the cave. On the surface, we assume that blindness
 69 is strongly selected against, and Q is dictated by mutation-selection balance.

70 Calculating the frequency of the blindness allele

71 Within the cave, the life cycle is as follows. (1) Embryos become juveniles and experience constant,
 72 directional selection with relative fitnesses of $w_{bb} = 1 + s$, $w_{Bb} = 1 + hs$, and $w_{BB} = 1$, where $s \geq 0$
 73 and $h \in [0, 1]$. (2) Juveniles migrate into and out of the cave such that a fraction m of adults come from
 74 the surface and $1 - m$ from the cave, where $0 \leq m \leq 1$. (3) Adults generate gametes with one-way
 75 mutation, where $0 \leq u \leq 1$ is the probability that a functional B allele becomes a non-functional b allele.
 76 (4) Gametes unite randomly to produce embryos. Given this life cycle, we calculate the allele frequency of
 77 the daughter generation (q') via standard equations:

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

$$q_a = q_j(1 - m) + Qm \quad \text{immigration} \quad (1b)$$

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

78 **Analysis of the change in allele frequency.** The change in allele frequency in one generation is $\Delta q =$
 79 $q' - q$, and each parameters influences it differently. Selection and mutation are directional forces, and
 80 increasing s or u increases Δq for $0 \leq q \leq 1$. (The derivatives are non-negative.) Increasing h causes
 81 selection to be more effective at low q , as rare b alleles are exposed to selection, but less effective at high
 82 q , as rare B alleles are sheltered from selection. Increasing h increases Δq if $0 < q < \left(1 + \sqrt{1 + s}\right)^{-1}$ and
 83 decreases it if $\left(1 + \sqrt{1 + s}\right)^{-1} < q < 1$. (The derivative is positive below this threshold and negative
 84 above it.) Migration harmonizes the allele frequency in the cave to the surface allele frequency. Thus
 85 increasing m increases Δq for low q and decreases Δq for high q . (The derivative is positive only when
 86 $0 \leq q < q_z(h, s, Q) \leq Q$, where q_z is a function describing a threshold.) However, increasing Q increases
 87 Δq for $0 \leq q \leq 1$. (The derivative is non-negative.)

88 Identifying equilibrium allele frequencies

89 The model we have developed is an example of migration-selection balance (Wright, 1969; Hedrick, 2011;
90 Nagylaki, 1992), extended to also include mutation. An equilibrium exists for this model when $\Delta q = 0$. For
91 small s , there is only one equilibrium, and it is near 0. For large s , there is only one equilibrium, and it is near
92 1. Three equilibria will only exist for moderate levels of selection (Figure 1). If $s = m = u = 0$, all forces of
93 evolution are eliminated and $\Delta q = 0$ for $0 \leq q \leq 1$. A lower bound for any valid equilibrium is $\frac{mQ(1-u)+u}{m(1-u)+u}$
94 (Lemma 1). An upper bound for any equilibrium is $1 - m(1 - u)(1 - Q)$ (Lemma 2). Furthermore, it is
95 important to note that

$$Q \leq \frac{mQ(1-u)+u}{m(1-u)+u} \implies Q \leq \hat{q} \quad (2)$$

96 indicating that the equilibrium frequency in the cave will be greater than or equal to the allele frequency
97 on the surface.

98 Assuming $s > 0$, the solution to $\Delta q = 0$ are the roots of the following cubic polynomial

$$g(q) = Aq^3 + Bq^2 + Cq + D = 0 \quad (3)$$

99 where

$$A = -s(1 - 2h)$$

$$B = s(1 - m(1 - u)(1 - Q) - h(3 + u - m(1 - u)(1 - 2Q)))$$

$$C = -(m(1 - u) + u) + sh(1 + u - m(1 - u)(1 - 2Q))$$

$$D = Qm(1 - u) + u$$

100 There are three possible roots of this equation, corresponding to three possible equilibria. Depending
101 on the parameter values, Equation 3 may have three real roots or one real root and two imaginary roots.
102 While the values of the roots of this polynomial can be expressed analytically, these equations are too
103 complex to be helpful for understanding the system. For simplicity, we will let \hat{q} represent any possible
104 equilibrium, and $\hat{q}_a \leq \hat{q}_b \leq \hat{q}_c$, stand for the roots of Equation 3.

105 **Protected polymorphism.** Rather than tackling the equilibria directly, we first demonstrate that the
 106 cave has a protected polymorphism. A protected polymorphism exists if the allele frequency moves
 107 away from both fixation and extinction, i.e. $\Delta q > 0$ when $q = 0$ and $\Delta q < 0$ when $q = 1$. For $q = 0$,
 108 $\Delta q = Qm(1 - u) + u$ and $q = 0$ will be an equilibrium only if $Qm = 0$ and $u = 0$; otherwise $\Delta q > 0$ at
 109 $q = 0$. For $q = 1$, $\Delta q = -m(1 - Q)(1 - u)$ and $q = 1$ will be an equilibrium if $m = 0$, $Q = 1$, or $u = 1$;
 110 otherwise $\Delta q < 0$. Thus a protected polymorphism always exists except at the edge cases $Qm = u = 0$,
 111 $m = 0$, $u = 1$, and $Q = 1$. In biological terms, the cave population will be polymorphic despite directional
 112 selection for b if there is some immigration from the surface population and the surface population is
 113 polymorphic.

114 **Validity of equilibria.** An equilibrium is only valid in our model if it is real and between $[0, 1]$; otherwise,
 115 it is not biologically interpretable in this system. Because there is a protected polymorphism, there
 116 will be either 1 valid, stable equilibrium, or 3 valid equilibria in a stable-unstable-stable configuration,
 117 depending on the parameter values. While we have not exhaustively determined the parameter ranges
 118 under which there will be only one valid equilibrium, we have determined that if $h \geq 1/3$ or if $h < 1/3$
 119 and $sh > \frac{m(1-u)+u}{1+u-m(1-u)(1-2Q)}$, there will be only one valid equilibrium (Lemma 3).

120 We can also estimate the amount of selection required such that q is an equilibrium ($g(q) = 0$, Equation
 121 3):

$$s_q(m, h, u, Q) = \frac{m(1-u)(q-Q) - (1-q)u}{q(q - q(q + m(1-Q)(1-u)) - h(1-q)(m(1-2Q)(1-u) - (1-2q) - u) + q)} \quad (4)$$

122 This equation is not valid for all $m \in [0, 1]$. If the migration rate is low, $m < \frac{(1-q)u}{(q-Q)(1-u)}$, no level of
 123 selection will make q an equilibrium, as all equilibria will be greater than q . Similarly, if the migration rate
 124 is high,

$$m > \frac{(1-q)(h(1-2q+u)+q)}{(1-u)(h(1-q)(1-2Q)+q(1-Q))}$$

125 no level of selection will make q an equilibrium, as all equilibria will be less than q .

126 **Dynamics and the evolution of blindness.** The dynamics of the evolution of the cave population
 127 depend on the parameter values and the starting allele frequency, q_0 . If there is one equilibrium, then the
 128 frequency of b will evolve monotonically towards it, i.e. $q_t \rightarrow \hat{q}$ as $t \rightarrow \infty$. If there are three equilibria, then
 129 the frequency of b will evolve monotonically to \hat{q}_a if $q_0 < \hat{q}_b$ and to \hat{q}_c if $q_0 > \hat{q}_b$.

130 When the cave population is founded, its initial allele frequency will likely match the equilibrium frequency
 131 on the surface ($q_0 = Q$). Because $Q < \hat{q}$ (Equation 2), the allele frequency in the cave will increase due
 132 to selection until it reaches the lowest equilibrium, i.e. $q_\infty = \inf\{q : 0 \leq q \leq 1 \text{ and } \Delta q = 0\}$. Whether
 133 blindness evolves in the cave depends on whether $q_\infty \geq q^*$, where q^* is a researcher-chosen threshold
 134 for determining that the cave population is a “blind” population. For example, $q^* = 0.5$ would specify
 135 that the blindness allele is the majority allele, and $q^* = 0.99$ would determine that the blindness allele is
 136 approximately fixed. We can also focus on phenotypes, and let $a = q^2 + 2q(1 - q)h$ measure the average
 137 blind phenotype in the cave; then

$$a_\infty \geq a^* \implies q_\infty \geq \frac{\sqrt{a^*(1 - 2h) + h^2} - h}{1 - 2h}$$

138 We define s^* as the minimum level of selection required for cave population to become blind, given the
 139 other parameters, i.e.

$$s^* = \inf\{s : s > 0 \text{ and } q_\infty \geq q^* \gg Q\}$$

140 For simplicity, we will only consider values of q^* much higher than the surface allele frequency. If there
 141 is one equilibrium, $s^* = s_{q^*}(m, h, u, Q)$; however, if there are three equilibria, q_t will evolve to the lower
 142 equilibrium and $q_\infty \approx Q \neq q^*$ (typically). Thus selection needs to be strong enough such that there is
 143 only one equilibrium; therefore,

$$s^* \approx \inf\{s : s > 0 \text{ and } s \geq s_{q^*}(m, h, u, Q) \text{ and } \Delta(s, m, h, u, Q) < 0\}$$

144 where $\Delta(s, m, h, u, Q)$ is the discriminant of Equation 3. Figure 2 plots analytical solutions for s^* based on

145 different thresholds. When $m \gg u$, the ratio s^*/m is roughly constant such that if $q_\infty \geq q^*$ then

$$\frac{s^*}{m} \geq \max \left\{ \frac{q^* - Q}{q^*(1 - q^*)(q^* + h(1 - 2q^*))}, \frac{1 - 6Q}{h} + \frac{2Q - 2\sqrt{Q^2 + hQ(1 - 3h(1 - 3Q) - 6Q)}}{h^2} \right\} \quad (5)$$

146 See Appendix for derivation.

147 Recessive Blindness

148 In order to study the equilibria in more detail we limit subsequent work to a model where blindness is
 149 recessive ($h = 0$). As we have previously shown the effects of varying h , its impact on subsequent results
 150 can be inferred generally. First, we will simplify our model by assuming that $u \ll 1$ such that $1 - u \approx 1$
 151 and

$$\Delta q \propto sq^2 [1 - q - m(1 - Q)] + [Qm + u - q(m + u)] \quad (6)$$

152 **Weak-selection approximation.** If selection is weak, then an equilibrium exists near $q = Q$. We use a
 153 second-order Taylor series at $q = 0$ to determine the upper bound on s for the presence of three equilibria
 154 (i.e. when selection is so strong that an equilibrium near Q does not exist). The second-order series
 155 allows us to determine the lower two equilibrium points; although, this approximation is inaccurate as q
 156 increases. This approximation gives us

$$\Delta q \approx s(1 - m)q^2 - (m + u)q + mQ + u \quad (7)$$

157 after assuming that $1 - Q \approx 1$. This equation has two roots, which are the lowest two of three total
 158 equilibria,

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

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

159 These two roots exist only if

$$0 < \sqrt{(m+u)^2 - 4s(1-m)(mQ+u)} \implies s < \frac{(m+u)^2}{4(1-m)(mQ+u)} \quad (8)$$

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

161 The derivative of Equation 7 is $\frac{d\Delta q}{dq}(q) = 2s(1-m)q - (m+u)$, and a equilibrium will be stable if
 162 $-2 < \frac{d\Delta q}{dq}(\hat{q}) < 0$. From this, it can be easily shown that $\hat{q}_{a,1}$ is stable and $\hat{q}_{b,1}$ is unstable.

163 **Strong-selection approximation.** In order to determine the lower bound on s for the presence of three
 164 equilibria, we assume that selection is strong enough such that $u/s \approx 0$ and $Q/s \approx 0$. Therefore,

$$\Delta q \propto -q [q^2 - [1 - m(1 - Q)]q + m/s] \quad (9)$$

165 and the equilibria can be described as

$$\begin{aligned} \hat{q}_{a,2} &= 0 \\ \hat{q}_{b,2} &= \frac{1}{2} \left(1 - m(1 - Q) - \sqrt{[1 - m(1 - Q)]^2 - \frac{4m}{s}} \right) \\ \hat{q}_{c,2} &= \frac{1}{2} \left(1 - m(1 - Q) + \sqrt{[1 - m(1 - Q)]^2 - \frac{4m}{s}} \right) \end{aligned}$$

166 The latter two equilibria will exist only if

$$s > \frac{4m}{[1 - m(1 - Q)]^2}$$

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

168 The derivative of Equation 9 is $\frac{d\Delta q}{dq}(q) = -3q^2 + 2[1 - m(1 - Q)]q - m/s$, and it can be easily shown
 169 that $\hat{q}_{b,2}$ is unstable and $\hat{q}_{c,2}$ is stable.

170 **Validity of approximations.** By substituting $\hat{q}_{a,1}$ and $\hat{q}_{b,1}$ back into Equation 6, we obtain $\Delta q =$
 171 $-s\hat{q}^2 (\hat{q} - Qm)$. Thus, $\Delta q \leq 0$, which indicates that $\hat{q}_{a,1}$ overestimates \hat{q}_a and that $\hat{q}_{b,1}$ underestimates \hat{q}_b .
 172 By substituting $\hat{q}_{b,2}$ and $\hat{q}_{c,2}$ back into Equation 6, we find that $\Delta q = Qm + u(1 - \hat{q})$. Thus $\Delta q \geq 0$, which
 173 indicates that $\hat{q}_{b,2}$ overestimates \hat{q}_b and that $\hat{q}_{c,2}$ underestimates \hat{q}_c . However, the error in our approxima-
 174 tions is slight (Figure 3).

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

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

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

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

184 Case 2b, $\frac{4m}{[1-m(1-Q)]^2} < s < \frac{(m+u)^2}{4(1-m)(mQ+u)}$: All three equilibria exist; \hat{q}_a and \hat{q}_c are stable, while \hat{q}_b is
 185 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
 186 towards \hat{q}_c .

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

189 Furthermore if $q_0 = Q$, the selection-threshold for blindness to be established in the cave is

$$s^* \approx \max \left\{ \frac{m(q^* - Q) - u(1 - q^*)}{q^{*2} (1 - q^* - m(1 - Q))}, \frac{(m + u)^2}{4(1 - m)(mQ + u)} \right\} \quad (10)$$

190 where q^* is the allele-frequency threshold.

191 **Finite-population simulations**

192 **Constant migration.** To investigate the impact of drift on our recessive-blindness model, we simulated
193 diploid populations of size $N = 1000$ by modifying our life cycle (Equation 1) to include a finite population:

194

$$q_j = \frac{(1+s)q^2 + q(1-q)}{(1+s)q^2 + (1-q^2)} \quad \text{selection} \quad (11a)$$

$$q_m = q_j(1-m) + Qm \quad \text{immigration} \quad (11b)$$

$$q_a \sim \text{Binomial}(q_m, 2N)/2N \quad \text{drift} \quad (11c)$$

$$q' = q_a + (1-q_a)u \quad \text{mutation} \quad (11d)$$

195 Here the adult population consists of $2N$ alleles sampled with-replacement from the post-immigration
196 gene pool.

197 For every simulation, $u = 10^{-6}$, $Q = 0.01$, and $q_0 = Q$. We varied s from 10^{-6} to 10^2 and m from 10^{-8}
198 to 1. We simulated 100 replicates for each combination of parameters; simulations were conducted for
199 10,000 or 5,000,000 generations. For each set of parameters, we recorded the average q' frequency across
200 these 100 populations at specific time points.

201 Our simulation results for finite populations are qualitatively similar to our analytical results for infinite
202 populations. For high migration rates, the average allele frequency is similar to the infinite model, except
203 that drift allows some populations that have three equilibria to evolve blindness (Figure 4B). However, at
204 low migration rates ($m < u/Q = 10^{-4}$), populations have low average frequency of b at 10,000 generations,
205 unless $s > 1$. As immigration decreased, these populations became dependent on *de novo* mutations
206 to produce b , which is a slow process. At 5 million generations, which is close to the estimated age of
207 cavefish populations (Gross, 2012), the average allele frequency is a better match to the results from the
208 the infinite-population model (Figure 4C); although, selection is ineffective for $s < 1/2N = 5 \times 10^{-4}$.

209 **Episodic migration.** Because cave and surface populations may be connected intermittently due to
210 flooding, we simulated periods of immigration followed by periods of isolation following a first-order

211 Markov process. The probability of switching between from isolation to immigration or vice versa was 10%
212 in each generation. Results for the intermittently connected simulations were nearly identical to previous
213 simulations, with the exception that at high levels of migration and selection, drift was more effective in
214 increasing average allele frequencies (Figure 4D).

215 Discussion

216 The evolution of blindness in caves has been hypothesized to result from relaxed selection and mutation
217 pressure and/or positive selection for alleles that result in eye loss. However, the degree to which these
218 factors interact and the theoretical level of selection required to induce blindness have not been quantified
219 previously. Here we show that in case of low level immigration into a cave, blindness will eventually
220 evolve, due to mutation and immigration of a few blindness alleles. This result fits the suggestion of some
221 previous hypotheses: relaxed selection can result in blind populations. However, for blindness to occur in
222 these conditions requires a significant amount of time. It is more likely that selection is much stronger
223 than previously anticipated, allowing blindness to be produced in caves over a relatively short period of
224 time. Furthermore, if levels of immigration are moderate to high, strong selection is necessary to produce
225 blind populations regardless of time.

226 Interestingly, although cave populations are likely small, drift is only essential to the evolution of blindness
227 in the cave population in a limited range of combinations of selection and migration for which we find
228 three equilibria. When immigration is low, low levels of selection can lead to blindness (lower left of Figure
229 4A); however, in finite populations stronger selection is required to overcome the effects of drift (lower left
230 of Figure 4C).

231 The amount of selection required for blindness to evolve depends on the migration rate and the level of
232 dominance of the blindness allele (Figure 2). For example, if $Q = 0.01$ and $h = 0$, the amount of selection
233 needs to be about 25 times the migration rate for a blind allele to become the major allele. Conversely, if
234 $h > 1/3$, it only needs to be about 3 times. The situation is reversed when we look at fixation. If $h = 0$,
235 selection needs to be about 100 times the migration rate for the frequency of the blind allele to exceed
236 99% in the cave. And if $h = 1$, it needs to be 10,000 times greater than the migration rate. If we focus on

237 phenotypes instead, we see that dominant alleles need lower levels of positive-selection to impact the
238 population (Figure 2).

239 The magnitude of selection coefficients required by our model to produce blindness given modest levels
240 of immigration are comparable to observations in many species. Levels of selection sufficient to produce
241 selective sweeps in wild populations range from 0.02–0.7 (Sáez et al., 2003; Schlenke and Begun, 2004;
242 Wootton et al., 2002; Nair et al., 2003). Estimated selection coefficients for drug resistance in *Plasmodium*
243 *falciparum* were 0.1–0.7, leading to fixation in 20–80 generations (Wootton et al., 2002; Nair et al., 2003).
244 For a major advantageous allele, the average value of s has been estimated as 0.11 in plants and 0.13 in
245 animals (Rieseberg and Burke, 2001; Morjan and Rieseberg, 2004). Recent work has suggested that eye
246 development imposes a high metabolic cost, particularly for juveniles (Moran et al., 2015). In a food-limited
247 environment this cost could lead to strong selection, but the precise degree of this selection is unknown.
248 The well-studied three-spine stickleback (*Gasterosteus aculeatus*) exhibits similar strong selection in a
249 novel environment. In experiments isolating armored sticklebacks in freshwater pools, armor was lost
250 within a few generations due to relaxed selection for defense and positive selection for the lower cost of
251 development in unarmored fish (Barrett et al., 2008). Estimates of selection in this species have ranged
252 from 0.13–0.16 (Terekhanova et al., 2014).

253 The selection coefficient of a blindness allele is determined not only by the amount of energy saved by
254 not having a visual system but also by any other pleiotropic effects, such as enhancement to feeding
255 ability (Jeffery, 2005). If an allele produces multiple, adaptive phenotypes, its selection coefficient is
256 more likely to be high enough to promote local adaptation and differentiation between cave and surface
257 populations. Genotype-dependent dispersal (Edelaar and Bolnick, 2012; Bolnick and Otto, 2013) is one
258 possible pleiotropic effect of blindness mutations that has not been considered in recent research on
259 cavefish. Ninety years ago, Lankester (1925) proposed that blindness evolves in caves because fish with
260 eyes may be attracted to sources of light and preferentially leave caves. In our model, emigration of
261 sighted individuals would be equivalent to increasing the selection coefficient, s , because individuals with
262 B alleles would systematically leave the cave. Even a low level of preferential emigration, e.g. 1%, would
263 provide a significant boost to local adaptation and the evolution of blindness in caves. It is quite possible
264 that in some species genotype-dependent dispersal combined with lower development costs promotes
265 the elimination of sight in caves despite the immigration of sightedness alleles from the surface.

266 While we have drawn conclusions about a single locus, multiple genes are involved in eye development
267 and sight. Loss-of-function mutations to any of these genes could produce blindness in caves. Linked
268 genes would effect our model by increasing the effective mutation rate of a sighted haplotype to a blind
269 haplotype, reducing the amount of selection required for the evolution of blindness. Unlinked genes
270 would provide more opportunities for drift to assist the evolution of blindness in caves.

271 We conclude that in most cases strong selection is necessary for the evolution of blind populations in caves.
272 This result is consistent with two different observations of cavefish: (1) phototactic fish may leave caves,
273 effectively selecting for the maintenance of mostly blind fish, and (2) the metabolic cost of eyes is very high.
274 Additionally, the model and results presented in this paper are applicable beyond the evolution of cave
275 populations, expanding existing migration-selection balance theory. We have developed approximations
276 that allow us to understand the interaction of selection, migration, and mutation. Through simulation we
277 have incorporated genetic drift into the model and determined that in some situations it can enhance the
278 power of selection to drive local adaptation. Periods of isolation can also be important in these situations.

279 **Acknowledgments**

280 This work was supported by Arizona State University's School of Life Sciences and Barrett Honors College.
281 Steven Wu, David Winter, Kael Dai, Michael Rosenberg, and Phil Hedrick provided helpful feedback on this
282 manuscript.

283 Appendix

284 All the proofs below were validated in Mathematica (Wolfram Research, Inc., 2015).

285 **Lemma 1.** *If $m > 0$ or $u > 0$, $\frac{mQ(1-u)+u}{m(1-u)+u}$ is a possible equilibrium, and there is no equilibrium less than it. If*
 286 *$m = u = 0$, 0 is an equilibrium.*

287 *Proof.* Case 1. Let $f(q) = q' - q$ represent the change in allele frequency over one generation (Equation 1).
 288 Let $\tilde{q} = \frac{mQ(1-u)+u}{m(1-u)+u}$. If $s = 0$ and $m > 0$ (or $u > 0$), $f(\tilde{q}) = 0$, and therefore \tilde{q} is an equilibrium for these
 289 parameters. Furthermore, if $s \geq 0$, $f(q) > 0 \forall q \in [0, \tilde{q})$. Therefore, there is no equilibrium lower than \tilde{q} .

290 Case 2. Let $m = u = 0$, $f(0) = 0$. □

291 **Lemma 2.** *$1 - m(1 - u)(1 - Q)$ is a possible equilibrium, and there is no equilibrium greater than it.*

292 *Proof.* Let $\tilde{q} = 1 - m(1 - u)(1 - Q)$ and $h = 0$. Since $\lim_{s \rightarrow \infty} f(\tilde{q}) = 0$, \tilde{q} is a potential equilibrium.
 293 Furthermore, if $0 \leq h \leq 1$ and $s \geq 0$, $f(q) < 0 \forall q \in (\tilde{q}, 1]$. Therefore, there is no equilibrium higher than
 294 \tilde{q} . □

295 The derivation of a tighter upper bound can be achieved by not assuming $h = 0$; however, we do not
 296 report it at this time.

297 **Lemma 3.** *Let $s > 0$. Let $m > 0$ or $u > 0$. If $h \geq 1/3$ or if $h < 1/3$ and $sh > \frac{m(1-u)+u}{1+u-m(1-u)(1-2Q)}$, $g(q)$
 298 (Equation 3) has exactly one root in $[0, 1]$.*

299 *Proof.* Let $m > 0$ or $u > 0$. Then $g(1) < g(0)$ and $g(1) \leq 0 \leq g(0)$. By the intermediate value theorem
 300 there is at least one root in $[0, 1]$. Let $s > 0$ and we will show that there is exactly one root for several
 301 cases.

302 Case 1. Let $1/2 < h \leq 1$. Then $g(-\infty) < 0$ and $g(\infty) > 0$. By the intermediate value theorem, $g(0)$ has
 303 at least one root below 0, between 0 and 1, and above 1. Since $g(0)$ is cubic, it can have at most 3 roots;
 304 therefore, there is exactly one root in $[0, 1]$.

305 Case 2. Let $h = 1/2$. $g(q)$ reduces to a quadratic equation with one root less than 0 and exactly one root
 306 in $[0, 1]$.

307 Case 3. Let $1/3 \leq h < 1/2$. $\frac{d^2g(q)}{dq^2} \leq 0$, and $g(q)$ is concave in $[0, 1]$. Thus $g(q)$ has exactly one root in
 308 $[0, 1]$.

309 Case 4. Let $0 \leq h < 1/3$ and $sh > \frac{m(1-u)+u}{1+u-m(1-u)(1-2Q)}$. Then $\frac{dg(q)}{dq}(-\infty) < 0$, $\frac{dg(q)}{dq}(0) \geq 0$, $\frac{dg(q)}{dq}(1) \leq 0$,
 310 and $\frac{dg(q)}{dq}(0) > \frac{dg(q)}{dq}(1)$. By the intermediate value theorem, there must be a local minimum in $(-\infty, 0]$
 311 and a local maximum in $[0, 1]$. Thus $g(q)$ has exactly one root in $[0, 1]$. \square

312 **Derivation of Equation 5.** In order to derive Equation 5 we first assume that $u = 0$. Then

$$\lim_{m \rightarrow 0} \frac{s_{q^*}(m, h, u, Q)}{m} = \frac{q^* - Q}{q^*(1 - q^*)(q^* + h(1 - 2q^*))}$$

313 However, we also need to determine when Δq has only one root. First we approximate Δq by a second-
 314 order Taylor series near $q = 0$.

$$\Delta q \approx s(1 - m)(1 - h(3 + 2hs))q^2 + (hs(1 - m) - m)q + mQ$$

315 Next we find

$$\lim_{m \rightarrow 0} \frac{\inf \{s : \Delta(s, m, h, u, Q) < 0\}}{m} = \frac{1 - 6Q}{h} + \frac{2Q - 2\sqrt{Q^2 + hQ(1 - 3h(1 - 3Q) - 6Q)}}{h^2}$$

316 where $\Delta(s, m, h, u, Q)$ is the discriminant of the Taylor approximation.

317 Equation 5 is the maximum of these two values.

318 References

- 319 Avise, J. C. and R. K. Selander, 1972. Evolutionary genetics of cave-dwelling fishes of the genus *Astyanax*.
 320 Evolution 26:1–19.
- 321 Barr, T. C., 1968. Cave ecology and the evolution of troglobites. chap. 2, Pp. 35–102, in T. Dobzhansky,
 322 M. Hecht, and W. Steere, eds. Evolutionary Biology, vol. 2. Springer, New York.

- 323 Barrett, R. D., S. M. Rogers, and D. Schluter, 2008. Natural selection on a major armor gene in threespine
324 stickleback. *Science* 322:255–257.
- 325 Bolnick, D. I. and S. P. Otto, 2013. The magnitude of local adaptation under genotype-dependent dispersal.
326 *Ecology and evolution* 3:4722–4735.
- 327 Borowsky, R., 2015. Regressive evolution: Testing hypotheses of selection and drift. chap. 5, Pp. 93–109, *in*
328 A. Keene, M. Yoshizawa, and S. McGaugh, eds. *Biology and Evolution of the Mexican Cavefish*. Academic
329 Press, San Diego, CA.
- 330 Bradic, M., P. Beerli, F. J. García-de León, S. Esquivel-Bobadilla, and R. L. Borowsky, 2012. Gene flow and
331 population structure in the mexican blind cavefish complex (*Astyanax mexicanus*). *BMC Evol. Biol.* 12:9.
- 332 Bulmer, M., 1972. Multiple niche polymorphism. *American Naturalist* Pp. 254–257.
- 333 Coghill, L. M., C. D. Hulsey, J. Chaves-Campos, F. J. G. de Leon, and S. G. Johnson, 2014. Next generation
334 phylogeography of cave and surface *Astyanax mexicanus*. *Mol. Phylogenet. Evol.* 79:368–374.
- 335 Culver, D. C., L. L. Master, M. C. Christman, and H. H. Hobbs, 2000. Obligate cave fauna of the 48 contiguous
336 United States. *Conserv. Biol.* 14:386–401.
- 337 Darwin, C., 1859. *On the origin of species based on natural selection, or the preservation of favoured races*
338 *in the struggle for life*. 6th ed. John Murray, London.
- 339 Dowling, T. E., D. P. Martasian, and W. R. Jeffery, 2002. Evidence for multiple genetic forms with similar
340 eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. *Mol. Biol. Evol.* 19:446–455.
- 341 Edelaar, P. and D. I. Bolnick, 2012. Non-random gene flow: an underappreciated force in evolution and
342 ecology. *Trends in ecology & evolution* 27:659–665.
- 343 Gross, J. B., 2012. The complex origin of *Astyanax* cavefish. *BMC Evol. Biol.* 12:105.
- 344 Gross, J. B., R. Borowsky, and C. J. Tabin, 2009. A novel role for Mc1r in the parallel evolution of depigmen-
345 tation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genet.* 5:e1000326.
- 346 Haldane, J. B. S., 1930. A mathematical theory of natural and artificial selection.(part vi, isolation.). *in*

- 347 Mathematical Proceedings of the Cambridge Philosophical Society, vol. 26, Pp. 220–230. Cambridge
348 Univ Press.
- 349 Hedrick, P. W., 2011. Gene flow and selection. chap. 7.IV, Pp. 424–436, *in* Genetics of Populations, 4th ed.
350 Jones and Bartlett, Sudbury, MA.
- 351 Hinaux, H., J. Poulain, C. Da Silva, C. Noirot, W. R. Jeffery, D. Casane, and S. Rétaux, 2013. De novo
352 sequencing of *Astyanax mexicanus* surface fish and Pachón cavefish transcriptomes reveals enrichment
353 of mutations in cavefish putative eye genes. PloS One 8:e53553.
- 354 Jeffery, W. R., 2001. Cavefish as a model system in evolutionary developmental biology. Dev. Biol. 231:1–12.
355 ———, 2005. Adaptive evolution of eye degeneration in the Mexican blind cavefish. J. Hered. 96:185–196.
- 356 Kimura, M. and J. L. King, 1979. Fixation of a deleterious allele at one of two” duplicate” loci by mutation
357 pressure and random drift. Proceedings of the National Academy of Sciences 76:2858–2861.
- 358 Lahti, D. C., N. A. Johnson, B. C. Ajie, S. P. Otto, A. P. Hendry, D. T. Blumstein, R. G. Coss, K. Donohue, and
359 S. A. Foster, 2009. Relaxed selection in the wild. Trends in ecology & evolution 24:487–496.
- 360 Lankester, E., 1925. The blindness of cave-animals. Nature 116:745–746.
- 361 Moran, D., R. Softley, and E. J. Warrant, 2015. The energetic cost of vision and the evolution of eyeless
362 Mexican cavefish. Science Advances 1:e1500363.
- 363 Morjan, C. L. and L. H. Rieseberg, 2004. How species evolve collectively: implications of gene flow and
364 selection for the spread of advantageous alleles. Mol. Ecol. 13:1341–1356.
- 365 Nagylaki, T., 1992. Migration and selection. chap. 6, Pp. 128–152, *in* Introduction to Theoretical Population
366 Genetics. Springer-Verlag, Berlin.
- 367 Nair, S., J. T. Williams, A. Brockman, L. Paiphun, M. Mayxay, P. N. Newton, J.-P. Guthmann, F. M. Smithuis,
368 T. T. Hien, N. J. White, et al., 2003. A selective sweep driven by pyrimethamine treatment in southeast
369 Asian malaria parasites. Mol. Biol. Evol. 20:1526–1536.
- 370 Niven, J. E., 2008. Evolution: convergent eye losses in fishy circumstances. Curr. Biol. 18:R27–R29.

- 371 Niven, J. E. and S. B. Laughlin, 2008. Energy limitation as a selective pressure on the evolution of sensory
372 systems. *J. Exp. Biol.* 211:1792–1804.
- 373 Protas, M., M. Conrad, J. B. Gross, C. Tabin, and R. Borowsky, 2007. Regressive evolution in the Mexican
374 cave tetra, *Astyanax mexicanus*. *Curr. Biol.* 17:452–454.
- 375 Protas, M. E., C. Hersey, D. Kochanek, Y. Zhou, H. Wilkens, W. R. Jeffery, L. I. Zon, R. Borowsky, and C. J.
376 Tabin, 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism.
377 *Nat. Genet.* 38:107–111.
- 378 Rieseberg, L. H. and J. M. Burke, 2001. The biological reality of species: gene flow, selection, and collective
379 evolution. *Taxon* 50:47–67.
- 380 Rohner, N., D. F. Jarosz, J. E. Kowalko, M. Yoshizawa, W. R. Jeffery, R. L. Borowsky, S. Lindquist, and C. J.
381 Tabin, 2013. Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in
382 cavefish. *Science* 342:1372–1375.
- 383 Sáez, A. G., A. Tatarenkov, E. Barrio, N. H. Becerra, and F. J. Ayala, 2003. Patterns of DNA sequence
384 polymorphism at Sod vicinities in *Drosophila melanogaster*: unraveling the footprint of a recent selective
385 sweep. *Proc. Natl. Acad. Sci. U. S. A.* 100:1793–1798.
- 386 Schlenke, T. A. and D. J. Begun, 2004. Strong selective sweep associated with a transposon insertion in
387 *Drosophila simulans*. *Proc. Natl. Acad. Sci. U. S. A.* 101:1626–1631.
- 388 Strickler, A. G., Y. Yamamoto, and W. R. Jeffery, 2007. The lens controls cell survival in the retina: evidence
389 from the blind cavefish *Astyanax*. *Dev. Biol.* 311:512–523.
- 390 Terekhanova, N. V., M. D. Logacheva, A. A. Penin, T. V. Neretina, A. E. Barmintseva, G. A. Bazykin, A. S.
391 Kondrashov, and N. S. Mague, 2014. Fast evolution from precast bricks: genomics of young freshwater
392 populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genet.* .
- 393 Wilkens, H., 1988. Evolution and genetics of epigeal and cave *Astyanax fasciatus* (Characidae, Pisces). Pp.
394 271–367, in T. Dobzhansky, M. Hecht, and W. Steere, eds. *Evolutionary biology*. Springer.
- 395 Wolfram Research, Inc., 2015. *Mathematica* 10.1.

- 396 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
397 diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* 418:320–323.
- 398 Wright, S., 1931. Evolution in mendelian populations. *Genetics* 16:97.
- 399 ———, 1969. Selection and immigration. Pp. 36–38, *in* *Evolution and the Genetics of Populations*, vol. 2.
400 The University of Chicago Press, Chicago.
- 401 Yamamoto, Y., D. W. Stock, and W. R. Jeffery, 2004. Hedgehog signalling controls eye degeneration in blind
402 cavefish. *Nature* 431:844–847.
- 403 Yeaman, S. and S. P. Otto, 2011. Establishment and maintenance of adaptive genetic divergence under
404 migration, selection, and drift. *Evolution* 65:2123–2129.
- 405 Yeaman, S. and M. C. Whitlock, 2011. The genetic architecture of adaptation under migration-selection
406 balance. *Evolution* 65:1897–1911.

407 **Figures**

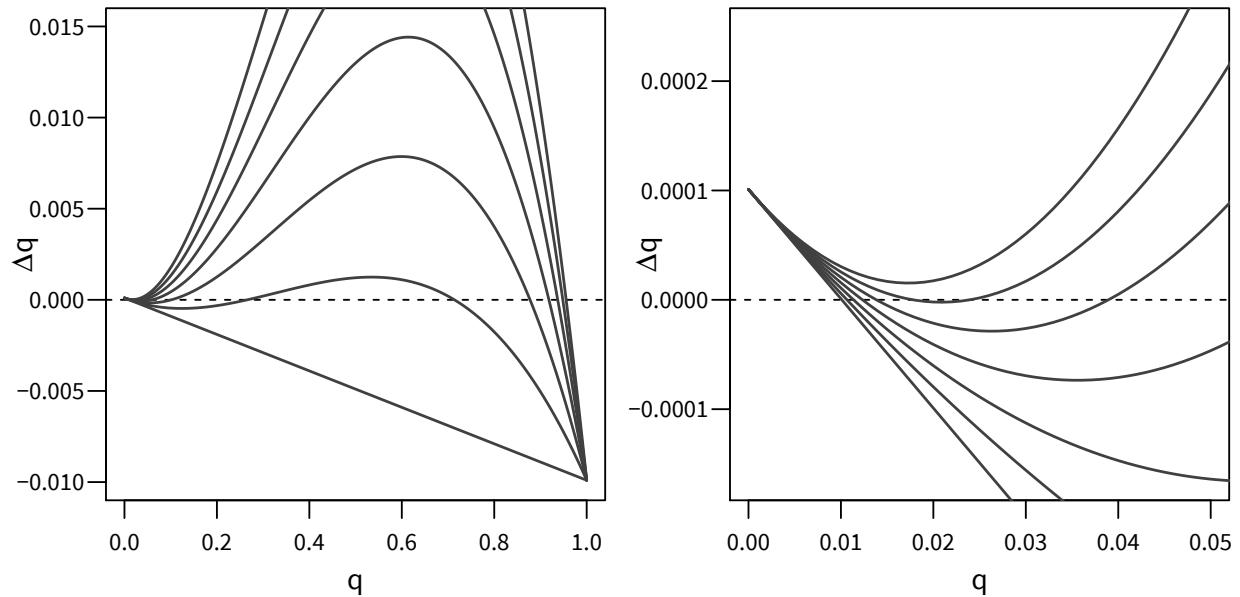


Figure 1: As selection increases, the evolutionary dynamics of the cave population changes. When s is low (bottom line; $s = 0$), there is only one equilibrium: near 0. As s increases (middle five lines, $s = 0.05, 0.1, 0.15, 0.2$, and 0.25) the local maximum (upper hump) increases and crosses the x-axis, 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 $m = 0.01$, $h = 0$, $u = 10^{-6}$, and $Q = 0.01$. The figure on the right is an enlarged view of a small part of the figure on the left.

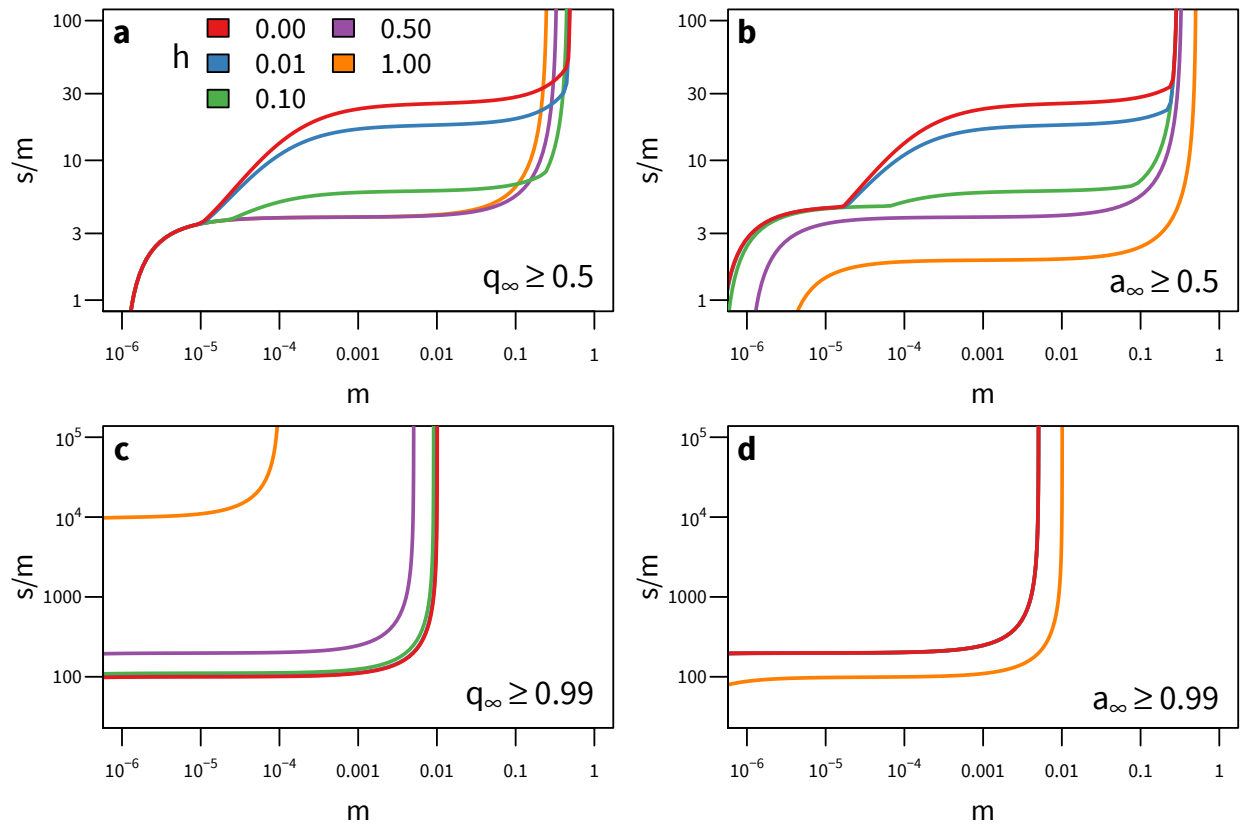


Figure 2: The level of dominance (h) of the blindness allele (q) affects the level of selection (s) required to produce blind populations. Each line represents how strong selection must be relative to migration (m) for blindness to evolve in the cave for a given level of dominance (s^*/m). Regions above the curves produce populations that are “blind” and regions below are not. Each panel contains a different condition for defining whether the cave is blind. (a) For the blind allele to become the majority allele requires stronger selection when the allele is recessive ($h = 0$). (b) For the blind phenotype to become the majority phenotype requires stronger selection when the allele is recessive. (c) For the blind allele to become fixed requires stronger selection when the allele is dominant. (d) For the blind phenotype to become fixed requires stronger selection when the allele is recessive. The curves were calculated analytically with $u = 10^{-6}$ and $Q = 0.01$.

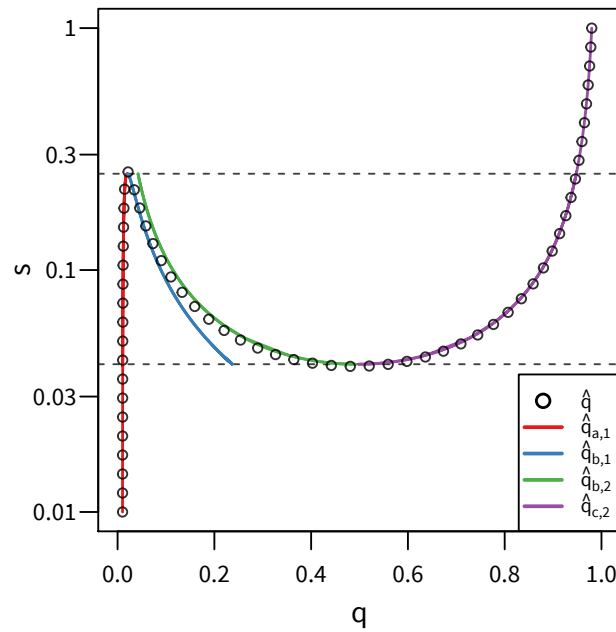


Figure 3: Our recessive-blindness equilibria approximations are accurate. The approximations developed in this paper (solid lines) are a good fit for calculated values of s that result in equilibrium for a given q (circles) using Equation 3. The dashed lines are our approximate bounds for the existence of three equilibria (i.e. for small and large values of s , there is one equilibrium; for intermediate values of s there are three possible equilibria). Other parameters are $m = 0.01$, $u = 10^{-6}$, and $Q = 0.01$.

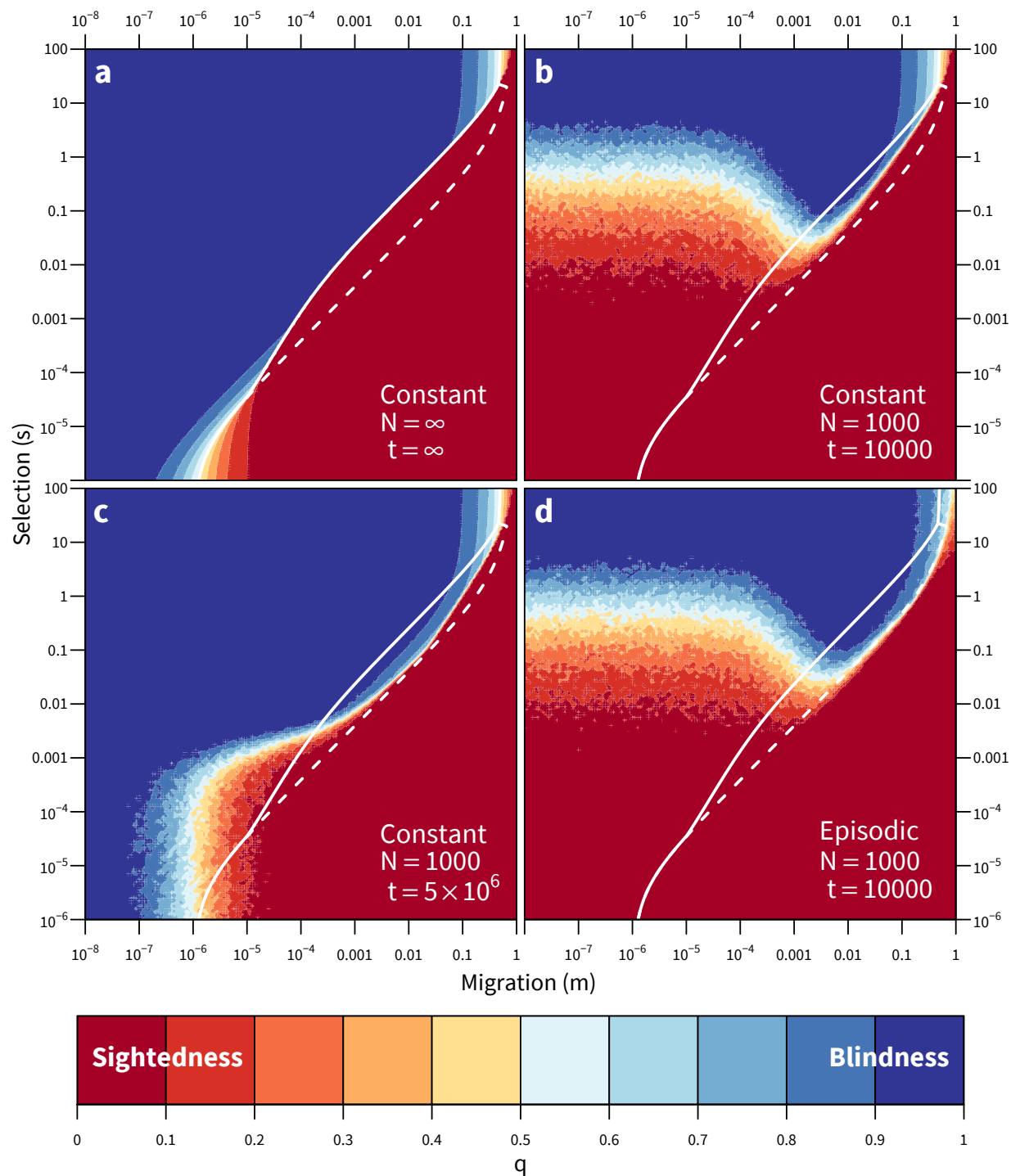


Figure 4

Figure 4: Populations evolve blindness in the face of immigration only with the help of strong selection. (a) The equilibrium frequency of the blindness allele (b) for an infinite population, and (a–d) average frequencies of b after t generations in finite populations (100 replicates) with either constant or episodic migration. Colors correspond to the frequency of the blindness allele (b) for a given combination of selection (s) and migration (m), where blue is high frequency (blindness evolved) and red is low (blindness did not evolve). The solid white line corresponds to $s_{0.5}^*$. The area between the solid and dashed lines corresponds to the region where three equilibria exist. Other parameters are $u = 10^{-6}$, $Q = 0.01$, and $q_0 = Q$.