Technical comment on 'Negative-assortative mating for color in wolves'

Background

Hedrick et al. (2016) reported on "negative-assortative mating for color in wolves" from Yellowstone National Park, the "first documented case of significant negative-assortative mating in mammals." Based on the close correspondence of genotype and allele frequencies observed in the wild to that predicted by their population genetic model, they conclude that "negative-assortative mating could be entirely responsible for the maintenance of this well-known color polymorphism." While researching examples of nonrandom mating in the wild to teach in class I discovered that the results of their population genetic model are inconsistent with their stated assumptions, as I understand them. In this paper, I revisit the model with the following two objectives:

1. Demonstrate that the frequency of negative-assortative mating between gray and black pelage color morphs in their model does not follow from their assumptions; and
2. Derive results that are consistent with their assumptions.

I am critiquing only their model, not the data analysis or conclusions. Both the original model and the new model analyzed here lead to similar inferences about the maintenance of the pelage color polymorphism because the equilibrium genotype and allele frequencies are nearly the same in both models. However, it is important that the mathematical biology literature provide logically consistent analysis so that future researchers may benefit most from its insights.

The frequency of assortative mating is inconsistent with the assumptions

Hedrick et al. (2016) assume that a proportion A matings are assortative, but the proportion they derive is much less (see Fig. 1 for a graphical derivation). For consistency, I use the same symbols as Hedrick et al. (2016) (Table 1 lists all symbols and their definitions). I infer three key assumptions from the two statements on the bottom-left of pg. 758 of Hedrick et al. (2016):

"gray wolves have a genotype of kk and an assumed frequency of P and black wolves have genotypes Kk and KK with frequencies H and Q, respectively ($P + H + Q = 1$)."

"Assume that A and $1 - A$ are the proportions of negative-assortative mating and random mating, respectively, in the population."

From these statements, I infer that:

1. kk, Kk, and KK are mutually exclusive genotypes with frequencies P, H, and Q
2. Negative-assortative mating and random mating are mutually exclusive mating types with frequencies A and $1 - A$

3. Genotype and mating type are independent $(\Pr[A \cap B] = \Pr[A] \times \Pr[B])$

Other standard population genetic assumptions such as infinite population size also apply.

Table 1: Glossary of mathematical symbols, variable string used in source code, and description.

Symbol	Variable string	Description
k	k	recessive beta-defensin variant
K	K	dominant beta-defensin variant
p	p	frequency of k allele
q	q	frequency of K allele
P	P	frequency of kk genotype
H	H	frequency of Kk genotype
Q	Q	frequency of KK genotype
A	A	proportion negative-assortatively mating

Based on these assumptions, I deduce that:

1. The probability of all genotype-mating type combinations must sum to 1
2. The probability of all genotypes in the assortative-mating subspace must sum to A
3. The probability of all genotypes in the random-mating subspace must sum to $1 - A$

There are six mutually exclusive genotype-mating type combinations in the population sample space (Fig. 1a). Since $P + H + Q = 1$, $A + 1 - A = 1$, and genotype is independent of mating type, the probability of all genotype-mating type combinations must sum to 1.

$$\sum_i^{\{kk,Kk,KK\}} \sum_j^{\{assort,random\}} \text{Genotype}_i \cap \text{Mating type}_j =$$

$$(kk \cap \text{assort}) + (Kk \cap \text{assort}) + (KK \cap \text{assort}) +$$
$$(kk \cap \text{random}) + (Kk \cap \text{random}) + (KK \cap \text{random})$$
$$= PA + HA + QA + P(1 - A) + H(1 - A) + Q(1 - A)$$
$$= A(P + H + Q) + (1 - A)(P + H + Q)$$
$$= 1 - A + A$$
$$= 1$$

Furthermore, we know that within the negative-assortative and random mating subspaces, the total probability must sum to A and $1 - A$, respectively:

$$A = PA + HA + QA$$
$$1 - A = P(1 - A) + H(1 - A) + Q(1 - A)$$

45 The model in Hedrick et al. (2016) is internally inconsistent because the proportion of negative-assortative
46 matings does not equal A as defined (Fig. 1b).

$$A := \frac{\text{Assortative matings}}{\text{Total matings}} = \frac{\text{Assortative matings}}{\text{Assortative} + \text{Random matings}}$$

47 Hedrick et al. (2016) state that the frequency of assortative matings is $2AP(H + Q)$ (*cf* top of pg. 759)
48 and the frequency of random matings is $1 - A$. Applying these frequencies reveals that the proportion of
49 assortative matings is not equal to A as assumed:

$$\frac{\text{Assortative matings}}{\text{Assortative} + \text{Random matings}} = \frac{2AP(H + Q)}{2AP(H + Q) + 1 - A}$$

50 There are no solutions to the expression above where the proportion of negative-assortative mating would
51 equal A when genotype and mating proportions are between 0 and 1. With their model, the actual proportion
52 of negative-assortative would vary from 0 when $P = 0$ or $P = 1$ and $A/(2 - A)$ when $P = 0.5$.

53 To summarize, a proportion A should mate assortatively given the assumptions of their model, but only
54 $2AP(H + Q)$ actually mate assortatively according to their results. In essence, they assign a proportion A
55 to mate assortatively, but then a proportion $A - 2AP(H + Q)$ do not mate assortatively (the area of the
56 gray regions in Fig. 1b), and are therefore not counted among the total number of matings. This is why the
57 probabilities of all matings do not sum to 1. The logically consistent solution is to condition on the fact that
58 if a mating is negative-assortative it must by definition have one gray and one black parent (Fig. 1b). In the
59 next section, I use this approach to derive new results. In contrast, Hedrick et al. (2016) effectively impose
60 selection without ever stating that assumption. To deal with the fact that resulting genotype frequencies do
61 not sum to 1, they regularize the frequencies (*cf* equation 1a-b), as normally done in models of selection.
62 Regularization is appropriate with selection because selection shrinks or expands the sample space as long as
63 average fitness does not equal 1. Hedrick et al. (2016) impose selection because some of the assortative-mating
64 individuals do not mate assortatively and are therefore not counted among the matings that result in offspring.

Revised solutions consistent with model assumptions

66 The previous section showed that the frequency of assortative gray × black matings was not derived in a
67 manner logically consistent with the model's assumptions. Here I derive new mating frequencies, genotype
68 frequencies, and equilibria. I used Sympy version 1.7.1 (Meurer et al. 2017) for symbolic derivations through
69 Python version 3.6 and the R package **reticulate** version 1.25 (Ushey et al. 2022). All other computations
70 were performed in R version 4.2.0 (R Core Team 2022). The source code is available in a public GitHub
71 repository and will be archived on Zenodo upon publication.

72 Table 2 derives the probabilities of all possible outcomes and Table 3 summarizes the frequency each mating
73 combination. This is the exact same process used to derive frequency of mating combinations in positive-
74 assortative mating models (e.g. Hedrick and Ritland 2012). Since I model random mating identically to the

previous model, the frequencies of gray × gray and black × black matings are identical; only the frequency of gray × black matings differs between models (Table 3; Fig. @ref{fig:sample-space}c).

Code in the Supporting Information derives the expressions in (Table 3) analytically using a computer algebra system, but one can also use the Law of Total Probability to prove it. The Law of Total Probability for discrete probability distributions states that $\Pr[A] = \sum_i \Pr[A|B_i]\Pr[B_i]$ where $\Pr[A|B_i]$ is the probability of outcome A conditional on outcome B_i. The total probability of A is the sum of conditional probabilities across all outcomes for event B. Using the Law of Total Probability, the probability of a gray × black mating is:

$$\Pr[\text{gray} \times \text{black}] = \Pr[\text{gray} \times \text{black}|\text{assort}]\Pr[\text{assort}] + \Pr[\text{gray} \times \text{black}|\text{random}]\Pr[\text{random}]$$

We already assume that $\Pr[\text{assort}] = A$ and $\Pr[\text{random}] = 1 - A$. With random mating, I arrive at the same expression as Hedrick et al. (2016), $\Pr[\text{gray} \times \text{black}|\text{random}] = 2P(H + Q)$ (*cf* top of pg. 759). If the mating is negative-assortative, then it *must* be a gray × black mating. Therefore, $\Pr[\text{gray} \times \text{black}|\text{assort}] = 1$. Putting these together, I obtain:

$$\begin{aligned}\Pr[\text{gray} \times \text{black}] &= 1 \times A + 2P(H + Q) \times (1 - A) \\ &= A + 2P(H + Q)(1 - A)\end{aligned}$$

This result diverges from that given in Hedrick et al. (2016), where they report the frequency of gray × black matings is $2P(H + Q)$ (*cf* Table 1).

Table 2: The probability of every mating outcome in the negative-assortative mating model analyzed by Hedrick *et al.* (2016). For the notation, the probability of event X is $\Pr[X]$. The total probabilities for each row are derived from the product of all probabilities in the same row, $\Pr[\text{Total}] = \Pr[\text{Parent 1}] \times \Pr[\text{Mating}] \times \Pr[\text{Parent 2}]$.

Parent 1	Pr[Parent 1]	Mating	Pr[Mating]	Parent 2	Pr[Parent 2]	Pr[Total]	Color
kk	P	assortative	A	kk	0	0	Gray × gray
kk	P	random	$1 - A$	kk	P	$P^2(1 - A)$	Gray × gray
kk	P	assortative	A	$K-$	1	AP	Gray × black
kk	P	random	$1 - A$	$K-$	$H + Q$	$P(H+Q)(1-A)$	Gray × black
$K-$	$H + Q$	assortative	A	kk	1	$A(H + Q)$	Gray × black
$K-$	$H + Q$	random	$1 - A$	kk	P	$P(H+Q)(1-A)$	Gray × black
$K-$	$H + Q$	assortative	A	$K-$	0	0	Black × black
$K-$	$H + Q$	random	$1 - A$	$K-$	$H + Q$	$(H+Q)^2(1-A)$	Black × black

Table 3: Hedrick *et al.* (2016) incorrectly derive the frequency of
gray × black. The corrected expressions are provided here.

Color	Mating Genotypes	Frequency (Hedrick *et al.* 2016)	Frequency (this paper)
Gray × gray	$kk \times kk$	$P^2(1-A)$	$P^2(1-A)$
Gray × black	$kk \times K-$	$2P(H+Q)$	$A + 2P(H+Q)(1-A)$
Black × black	$K- \times K-$	$(H+Q)^2(1-A)$	$(H+Q)^2(1-A)$

Despite the different frequency of gray × black matings resulting from each model, the equilibrium genotype frequencies are very similar. In both models, $\hat{P} = 0.5$, implying $0.5 = \hat{H} + \hat{Q}$. I find that $\hat{Q} = (A/2 - \sqrt{2(A+1)} + 1.5)/(1-A)$, which is close to the equilibrium values obtained in Hedrick et al. (2016) through recursion (Fig. 2). Next, I compared allele frequency change depicted in Figs. 3-4 of Hedrick et al. (2016) to that predicted with the new model. The effect of A on change in the frequency of the K allele is qualitatively similar, but much faster with the new model (Fig. 3). This is because the magnitude of allele frequency change far from the equilibrium is much greater with the new model (Fig. 4). As a result, Hedrick et al. (2016) overestimate how long it would take for K to reach equilibrium given observed levels of assortative mating ($A = 0.430$). They conclude that K would reach equilibrium at $\hat{q} = 0.278$ in 25 generations with $A = 0.430$. In the revised model, K would reach equilibrium at $\hat{q} = 0.271$ in only 15 generations with $A = 0.430$ and a starting allele frequency $q_0 = 0.01$. Hence, the revised model actually lends credence to their conclusion that negative-assortative mating may be a better explanation than heterozygote advantage for variation at the beta definsin locus.

In conclusion, the logical inconsistency in Hedrick et al. (2016) does not undermine their primary conclusion that negative-assortative mating by color may explain the distribution of genotype frequencies at the beta definsin locus in the Yellowstone population of wolves (*Canis lupus*). The new derivation here may prove useful to future research on negative-assortative mating.

Literature cited

Hedrick, P. W., and K. Ritland. 2012. Population genetics of the white-phased "Spirit" bear of British Columbia. Evolution 66:305–313.

Hedrick, P. W., D. W. Smith, and D. R. Stahler. 2016. Negative-assortative mating for color in wolves. Evolution 70:757–766.

Meurer, A., C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, Š. Roučka, A. Saboo, I. Fernando, S. Kulal, R. Cimrman, and A. Scopatz. 2017. SymPy: Symbolic computing in Python. PeerJ Computer Science 3:e103.

R Core Team. 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ushey, K., J. J. Allaire, and Y. Tang. 2022. Reticulate: Interface to 'Python'.

(a) **Sample space**
(must sum to 1)

Mating Type

Genotype		negative assortative A	random $1-A$
kk	P	$kk \cap \text{assort}$ PA	$kk \cap \text{random}$ $P(1-A)$
Kk	H	$Kk \cap \text{assort}$ HA	$Kk \cap \text{random}$ $H(1-A)$
KK	Q	$KK \cap \text{assort}$ QA	$KK \cap \text{random}$ $Q(1-A)$

Six mutually exclusive genotype-mating type combinations comprise the sample space.

Their probabilitiess must sum to 1.

This sample space can be broken down into assortative mating and random mating subspaces with total probability of *A* and *1 - A.*

(b) **Assortative mating subspace**
(must sum to A)

This study

Black parent

Gray parent	Kk $H/(H+Q)$	Kk $Q/(H+Q)$
kk 1	$\dfrac{AH}{H+Q}$	$\dfrac{AQ}{H+Q}$

(c) **Random mating subspace**
(must sum to $1-A$)

Parent 1

Parent 2	kk P	$K-$ $H+Q$
kk P	$P^2(1-A)$	$P(H+Q)(1-A)$
$K-$ $H+Q$		$(H+Q)^2(1-A)$

$P(H+Q)(1-A)$

Both studies model randomly mating individuals identically

Hedrick et al.

Parent 1

Parent 2	kk P	$K-$ $H+Q$
kk P		$AP(H+Q)$
$K-$ $H+Q$		

$AP(H+Q)$

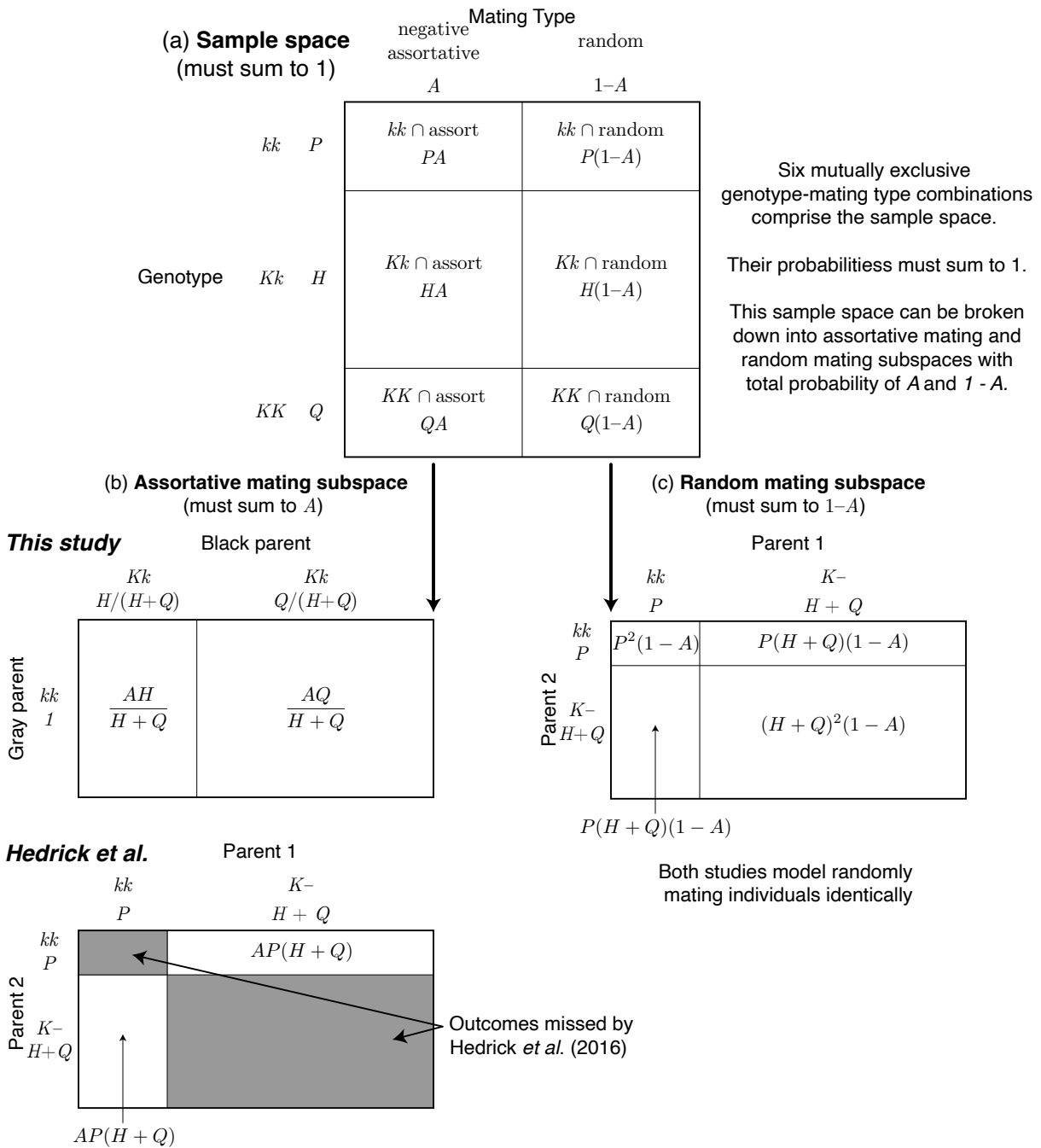Outcomes missed by Hedrick *et al.* (2016)

Figure 1: (Caption next page.)

6

Figure 1: (Previous page.) A graphical summary of the sample space that illustrates why this model arrives at results that differ from Hedrick *et al.* (2016). (a) The sample space consists of six mutually exclusive genotype-mating type combinations that must sum to 1 given the assumptions. (b) In this study we show that the the probabilities in the assortative mating subspace sum to A, as assumed, if one conditions on assortative mating consisting of a black and gray coat color parent. In contrast, Hedrick *et al.* (2016) effectively assume that a proportion of assortatively mating individuals equal to the area of the gray boxes do not contribute matings. This is why the models reach different conclusions. (c) Both models treat randomly mating individuals identically.
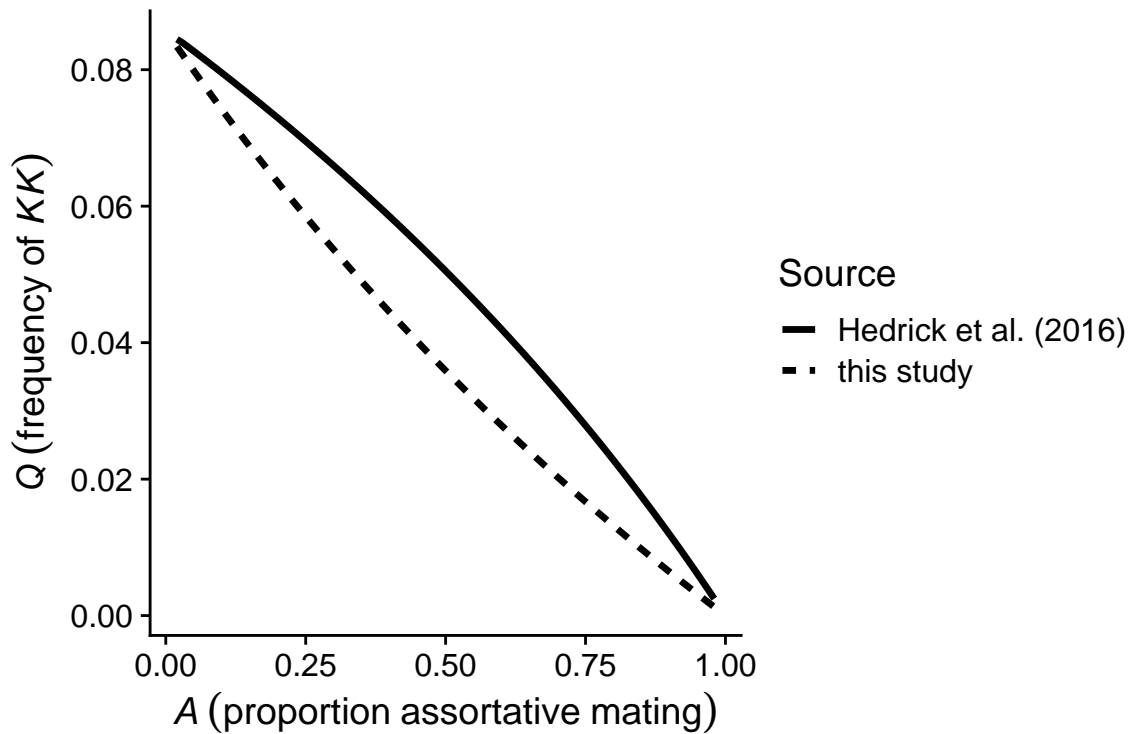

Figure 2: The equilibirum frequency of Q, the *KK* homozygote in this study (dashed line) and Hedrick *et al.* (2016) (solid line) for possible values of A, the proportion of wolves mating assortatively by color.
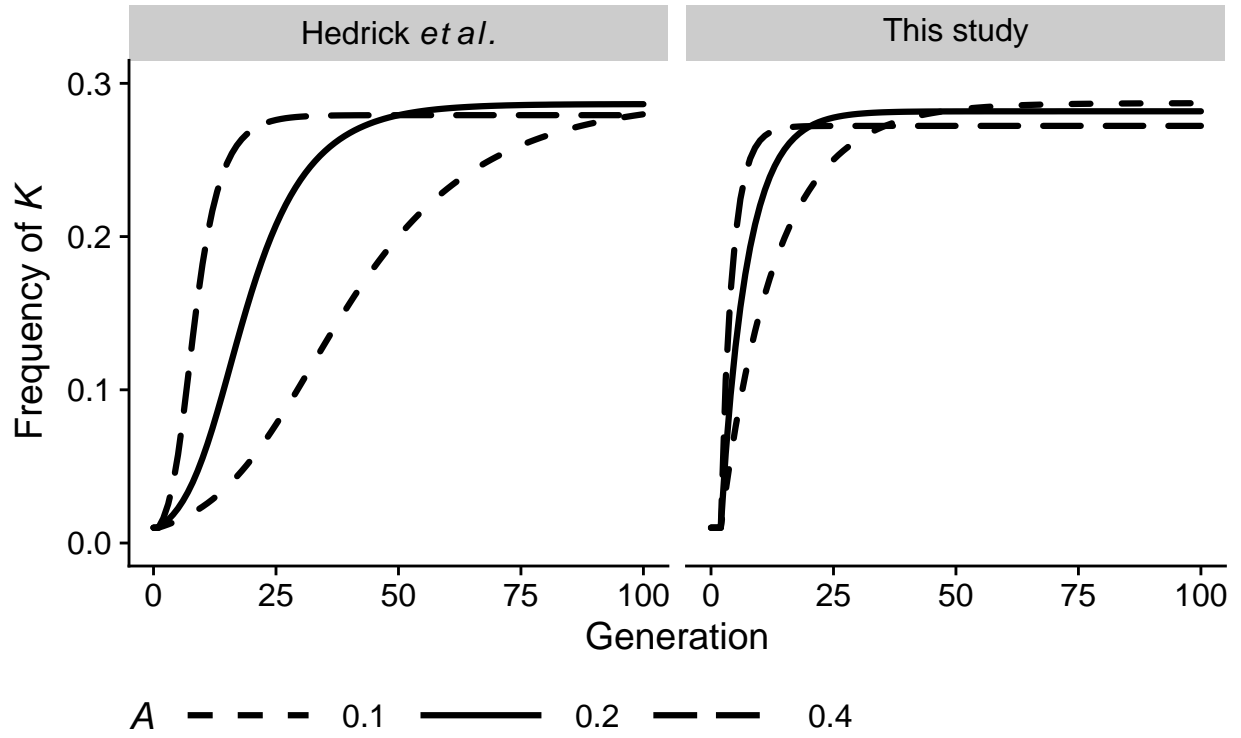
Figure 3: A comparison of the change in frequency of the black allele K between Hedrick *et al.* (2016) and this study. K starts at a frequency 0.01 and the plots show the dynamics for three different levels of negative-assortative mating A.
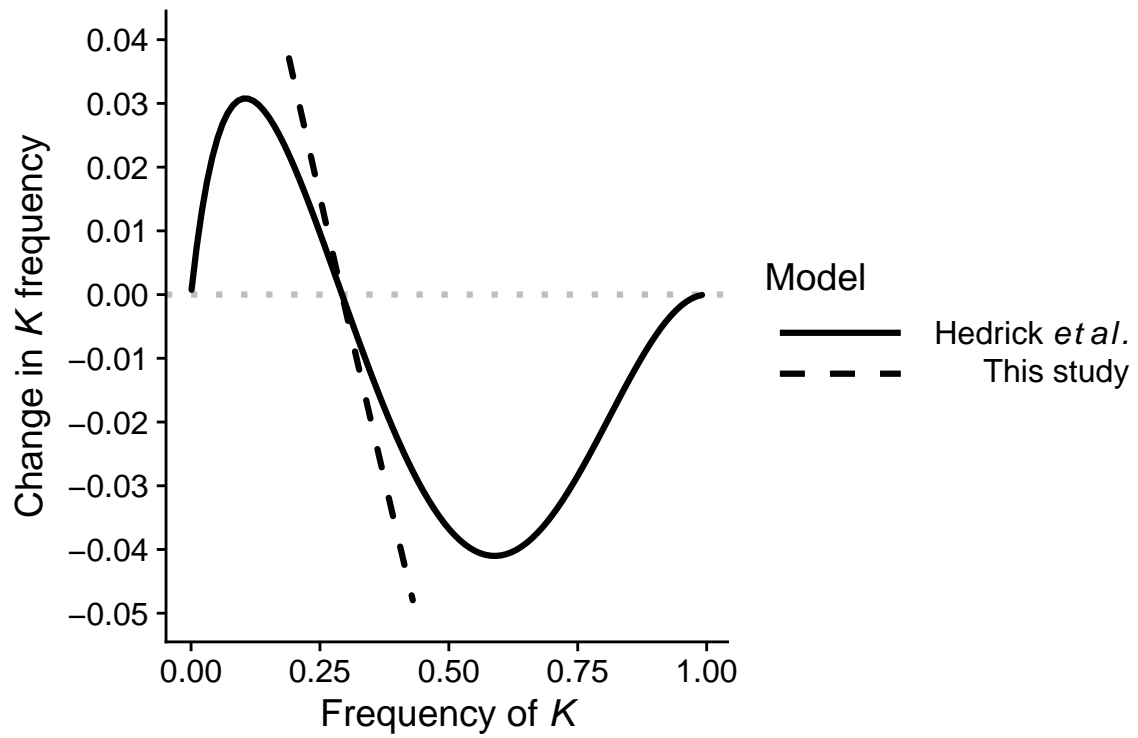
Figure 4: A comparison of the change in frequency of the black allele K between Hedrick *et al.* (2016) and this study. The initial frequency of K is along the x-axis and the change in K along the y-axis where $A = 0.430$ as in Hedrick *et al.* (2016) Fig. 4. The models have the similar equilibrua where they cross the gray dotted line at 0, but in this study the magnitude of change in allele frequency increases the system gets further from this equilibrium.