Continuous approximations for optimizing allele trajectories

Citation for published version:
Liu, AYH & Woolliams, JA 2010, 'Continuous approximations for optimizing allele trajectories' Genetics Research, vol. 92, no. 2, pp. 157-166.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
Genetics Research

Continuous Approximations for Optimising Allele Trajectories

by

A. Y. H. Liu and J. A. Woolliams

Roslin Institute, Midlothian EH25 9PS, United Kingdom

Short Title: Optimum Allele Trajectories

Key Phrases:

Continuous approximation, genotype assisted selection, GAS, QTL fixation, frequency trajectory

Proofs to be send to:

Ariel Yi-Hsuan Liu
Roslin Institute & Royal (Dick) School of Veterinary Science
Roslin BioCentre, Roslin, Midlothian EH25 9PS, United Kingdom
E-mail: ariel.liu@roslin.ed.ac.uk

Summary

The incorporation of genetic information such as quantitative trait loci (QTL) data into breeding schemes has become feasible as DNA technologies have advanced. Such strategies allow the frequency of desirable QTL to be controlled over a predefined time frame, allowing the allele trajectory for QTL to be manipulated. A continuous approximation to changes in allele frequency was developed to approximate the selection procedure as a continuous rather than a discrete process, and analytical solutions were obtained which shed light on how allele trajectories behave under different objective functions. Three different objectives were considered: (1) minimising the total selection intensity; (2) minimising the sum of squared selection intensities; and (3) equalising the selection intensity applied over time. Simulations and genetic algorithms were performed to test the accuracy and robustness of the continuous approximation. Theory shows that firstly the total selection intensity required for moving an allele from a starting frequency to another frequency point can be predicted independent of its trajectory, and secondly that objective (2) and (3) are equivalent as the number of selection opportunities (*T*) becomes large. The prediction of total selection intensity provides good fit for these two objectives, with the accuracy of prediction improving as *T* increases. However, for (1) the continuous approximation does not fit due to the existence of a discontinuous solution in which the continuous approximation is applied before the frequency of selected allele reaches 0.5 followed by rapid fixation.

1. INTRODUCTION

As identification of quantitative trait loci (QTL) becomes routine, genotype assisted selection (GAS) has become possible and even desirable for populations with managed breeding. GAS is where the frequency of a known allele, which affects the trait of selection, is managed generation by generation within the population, often to fixation. One of the known hurdles of the application of GAS is the Gibson effect, a phenomenon whereby GAS results in higher short-term genetic gain but lower long-term genetic gain than conventional selection methods which ignore the information on QTL (Gibson 1994). The explanation is, although GAS can fix the QTL in shorter time, the loss of variation on polygenes associated with the strong positive selection of QTL will lead to reduced selection response on polygenes, which can not be fully recovered. Various authors have shown that this effect can be ameliorated by optimising the selection procedures for the QTL over multiple generations, i.e. the optimisation of allele trajectory (Dekkers & van Arendonk 1998; Dekkers & Chakraborty 2001; Villanueva *et al.* 2002; Meuwissen & Sonesson 2004; Villanueva *et al.* 2004; Sanchez *et al.* 2006).

This optimisation process has led to a variety of approaches to manage the trajectory: maximising progress over the long term (Pong-Wong & Woolliams 1998; Villanueva *et al.* 2004), with pre-defined time horizons (Dekkers & van Arendonk 1998); or constrained to a constant rate of inbreeding (Villanueva *et al.* 2002). These studies make selection decisions based on estimated breeding values in one form or another, so optimum allele trajectory is therefore defined only implicitly. However the method of Dekkers & van Arendonk (1998) allows the optimised trajectory to be defined explicitly as a set of time points for the allele frequency, so defining the selection pressure that is directly applied to the allele to be analysed. Based on an observation of Dekkers & van Arendonk (1998), Meuwiseen & Sonesson (2004) directly defined the allele trajectory by making selection intensity on the allele constant over the period of selection. More recently, Sanchez et al (2006) pointed out that the effective population size was inversely proportional to the square of selection intensity, so that the optimum trajectory to minimise the accumulated inbreeding due to fixation should minimise the average squared selection intensity on the major gene over generations up to the given fixation time, in another words it should minimise the sum of squared selection intensities (simplified as sum of squared intensities thereafter) applied to the allele over generations.

A common theme to these studies of allele trajectories is that they use discrete generation models. This discrete time model imposes limitations on obtaining analytical solutions for the problem of approximating the optimal pathway, and the lack of analytical solution leaves unresolved the degree to which these approaches are distinct. Furthermore, the iterative solutions from the equalising selection intensities method leave open questions such as what is the total selection intensity (simplified as total intensity thereafter) required to fix the QTL given the circumstances. Therefore, this study establishes a continuous time model of the process of fixing an allele, and explicitly optimises the trajectory with respect to various objective functions of the selection intensity applied to the gene using the calculus of variations. This continuous model serves as a common platform allowing further investigation and comparison between different optimising objectives. The predictions from the continuous time model are compared to optimisations using discrete generations to quantify the precision of the continuous time model.

2. METHOD

(i) Theory

Continuous Approximations

Consider the process of moving a desired allele Q from frequency p_0 at time 0 to p_T at time T in discrete generations and assume for simplicity of notation that there is only one other allele, q, at that locus. The trajectory consists of the set of frequency points $\{p_t, t = 0, \ldots, T\}$ and optimisation of the trajectory is the set of p_t that maximise a certain objective function. Commonly, when considering fixation of alleles in GAS $p_0 = (2N)^{-1}$ and $p_T = 1$, as it models the fixing of a new mutation occurring in a diploid population of size N. This scenario is equivalent to the situation of eliminating a known allele from a population ($0 < p_0 < 1$ and $p_T = 0$), as removing one allele forces the frequency of all alternative alleles to 1. However the theory developed here will not be specific to these starting and finishing frequencies.

In this paper, following Meuwissen & Sonesson (2004) and Sanchez et al (2006), the objective functions considered are functions of the selection intensity applied directly to the allele. Let $p_{t,k}$ be the frequency of Q allele of individual k born at time t, so $p_{t,k}$ will take values 0, ½ or 1 depending on whether k has genotype qq, qQ or QQ. Using $p_{t,k}$ as the definition of an additive trait of selection, the population mean is p_t, the variance is $\frac{1}{2}p_t(1-p_t)$, the selection intensity i_t can be defined as:

$$\mathrm{i}_t = (p_{t+1} - p_t)/\sqrt{\tfrac{1}{2}\,p_t(1-p_t)} \qquad (1)$$

for $t = 0, \ldots, T-1$, and the trajectory is a sequence of points $\{p_t, t = 0, \ldots, T\}$.

The trajectory can be considered in continuous time rather than as a set of discrete generations. It is an assumption, to be tested later, that the use of continuous time will approximate the original problem better as the selection opportunities for changing allele frequency become greater, i.e. when *T* is large. Let the trajectory over time be given by $p(t)$, which is assumed to be a differentiable function of time *t*, then $\delta p = p_{t+\delta t} - p_t \approx p'(t)\delta t$ wher $p'(t) = \frac{dp}{dt}$,and $i_t \approx p'(t)\delta t/\sqrt{\tfrac{1}{2}\,p(t)(1-p(t))}$. Therefore provided the trajectory $p(t)$ is differentiable so that its derivative, $p'(t)$, exists, the sums over of the trajectory may be approximated by integrals.

Different objective functions that optimise the trajectory are considered and analysed using the continuous approximation, including: (1) the trajectory that minimises the total intensity; (2) the trajectory that minimises the sum of squared intensities; and (3) the trajectory that equalises selection intensity. Due to the amount of mathematical details involved, only the essential information and core equations are shown in this section, however, more details can be found in Appendix 1.

Minimising the Total Intensity

The total intensity for fixing an allele with a trajectory *p(t)* as *T* becomes large can be given by:

$$\sum_{t=0}^{T} i_t \approx \int_0^T \frac{p'(t)dt}{\sqrt{\frac{1}{2}p(t)(1-p(t))}} \qquad (2)$$

Transformation and integration of the above equation gives the following:

$$\sqrt{2}\,(sin^{-1}(1-2p_0) - sin^{-1}(1-2p_T)) \qquad (3)$$

Note that this solution only depends on the starting point p_0 and the ending point p_T, suggesting that there is no such thing as minimising the total intensity if the approximation is valid – the total intensity is fixed between a pair of frequency points regardless of its trajectory or the value of *T*. For a new mutation moving to fixation, $p_0 = (2N)^{-1}$ and $p_T = 1$, the total intensity applied to the allele during fixation is $\sqrt{2}\ [\frac{1}{2}\pi + sin^{-1}(1-N^{-1})]$, which tends to $\sqrt{2}\pi$ as *N* becomes large, i.e. when starting frequency approaches zero.

Minimising Sum of Squared Intensities

The specific optimisation considered by Sanchez et al (2006) was the trajectory of the allele frequencies required to minimise the impact of the process on accumulated inbreeding during the fixation. It is assumed here that the accumulated inbreeding can be well-approximated from summed rates of inbreeding (Δ*F*) achieved in each generation, and that the allele can be fully identified throughout the process so its frequency can be explicitly managed over time. The value of Δ*F* will vary according to the impacts of all the different selection advantages inherent in a selection scheme, not only the carrier status of individuals for the allele of interest (Woolliams & Bijma 2000), and will depend upon the square of the selection intensities applied (Woolliams *et al.* 1993). Therefore the objective for minimising the impact of the fixation is to minimise $\sum_{t=0}^{T} {i_t}^2$, which can be shown as follows:

$$\sum_{t=0}^{T} {i_t}^2 \approx \int_0^T \frac{p'(t)^2}{\frac{1}{2}p(t)(1-p(t))}dt \tag{4}$$

Solving the above equation gives $\sin^{-1}(1-2p) = At + B$ or, equivalently as:

$$p(t) = \tfrac{1}{2}[1 - \sin(At + B)] \tag{5}$$

, where *A* and *B* are constants of integration and vary depending on $p_{0,}$ p_T, and *T*. Values of *A* and *B* can be obtained by substituting these parameters into Equation (5). For example, assume that fixation is desired from a new mutation, i.e., $p_0 = (2N)^{-1} \approx 0$ for large *N*, and $p_T = 1$. With these conditions $B = \pi/2$ and $A = -\pi/\mathrm{T}$ to give $p(t) = \frac{1}{2}(1 - \sin[\frac{1}{2}\pi(1 - 2tT^{-1})])$. The optimal trajectory for minimising the sum of squared intensities applied to the allele is therefore a segment of a sine wave.

Equalising Selection Intensities

Based on the observation from Dekkers & van Arendonk (1998) that the selection intensities achieved in each generation are roughly constant in their simulated result with best long term gain, Meuwissen & Sonesson (2004) suggest optimising the trajectory to maximise the cumulative selection response is by making the selection intensities constant over time. Applying this objective in the continuous approximation gives a differential equation that is identical to that obtained above for the objective of minimising the sum of squared intensities. This indicates that the objective from Meuwissen & Sonesson gives an optimum trajectory identical to that from minimising the sum of squared intensities. This conclusion is analogous to the minimisation of sum of squares for n numbers whose sum is fixed to some value c − the solution has all numbers equal to c/n. Therefore the theory suggests that as the continuous approximation provides becomes more apt, so the distinction between the objectives of Sanchez et al (2006) and Meuwissen & Sonesson (2004) disappears. The question remains over how close an approximation.

(ii) Simulation Methods

Two types of simulation methods are included in this section, the first a genetic algorithm with small population size (N=10), and the other a simulation of breeding populations with large population size (N=500). Together they test the validity and robustness of the continuous approximation under various scenarios.

Genetic Algorithm

The genetic algorithm used differential evolution (Shepherd & Kinghorn 1992) to optimise the allele frequency in order to find the optimal trajectories with N=10, for the three objectives considered above: (i) equalising selection intensities; (ii) minimising sum of squared intensities; and (iii) minimising total intensity. Equalising the selection intensities was achieved by minimising the sum of all squared differences among the selection intensities.

Simulations of Breeding Schemes

Computer simulations of the breeding schemes start with a base population (t=0) of 500 diploid individuals and this population size was maintained through out the simulation.

One individual from the base population was randomly chosen to carry a single copy of positive allele (initial frequency $p_0 = (2N)^{-1}$) with allelic effect *a*, which equals 0.5 as the addition or removal of one positive allele result a change of 0.5 in terms of frequency. Random mating with possible selfing was assumed for simplicity, i.e. the genetic make-up of the offspring was randomly assigned from selected parents with replacement. As the theory shows that the objective of minimising sum of squared intensities resembles the objective of equalising selection intensities when *T* is large, only the objective of equalising selection intensities is used for its ease to execute. In addition, other selection strategies with oscillating intensities in a saw tooth pattern, i.e. intensity profiles of the form {0.3, 0.1, 0.3, 0.1…}, were also employed to test whether the continuous approximation still holds under more extreme conditions.

One should note that the time unit applied in this study was the opportunities for selection and mating. Hence the word cohort will be used hereafter to represent a group of animals which are the direct result of last selection and mating. The frequency of the positive allele was then calculated and recorded for each cohort, and simulation ended when the positive allele was either fixed ($p_t \geq (2N\text{-}1)/2N$) or lost ($p_t \leq (2N)^{-1}$). In the case of the allele being lost, the data was excluded from the final data set as we considered the pathway of allele fixation only. One thousand simulations were run for each set of parameters and the average number of cohorts required to fix selected allele was obtained to be compared to the expected number of cohorts required from the approximation.

Discrete generation: A pre-defined constant selection intensity (*i*) was applied over every cohort by restricting the average frequency of the selected individuals. Calculation was then performed for each cohort to obtain the target p_{t+1} from the p_t:

$$p_{t+1} = p_t + \ i\sqrt{\tfrac{1}{2}p_t(1-p_t)} \tag{6}$$

Selection candidates were composed of all individuals from the current cohort and were ranked according to their allelic value. Selection candidates were then removed sequentially from lower rank until the target p_{t+1} was achieved. However, as mating between selected parent are random, the average allele frequency in the resultant population could not be guaranteed and may deviate from the target p_{t+1}. For oscillating intensities, a similar procedure as described above was used, except that the intensity is not constant over every cohort.

Overlapping generation: The overlapping generation model was largely identical to the discrete generation model except that the candidates available for selection were not only restricted to the current cohort, but also extended to include 2 previous cohorts. For selection candidates with same allelic value, a randomisation process was used to determine which candidate would become a parent. Generation interval (*L*) was calculated as the age of parents (in units of cohorts) when the offspring born.

When several cohorts contribute to the selection, the genetic variance is higher than shown in Equation (6), i.e. $\frac{1}{2}p_t(1-p_t)$. Apart from the variance within all selected cohorts, the true genetic variance also contains an additional term for the variance between different cohorts:

$$V_{total} = \tfrac{1}{2}E[p](1 - E[p]) + \tfrac{1}{2}(E[p^2] - E[p]^2) \qquad (7)$$

where E[p] denotes expectations over the selected cohorts. Simulations were carried out using Equation (6) with V_{total} replacing $\frac{1}{2}p_t(1 - p_t)$. This was compared to using Equation (6) without modification.

3. RESULTS

As shown in the theory, the continuous approximation provides a prediction for the total intensity required to move a target allele from a specific frequency to another. The prediction is only affected by the starting and ending frequencies alone, and is independent of *T* or *N*, although in the case of new mutation, the starting frequency is inversely related to the population size. Assuming fixation is the goal (i.e. ending frequency is 1), the predicted total intensity is $\sqrt{2}\pi$ (≈ 4.44) for fixing a mutation in a large population, and 3.80 for a starting frequencies of 0.05.

(i) Goodness of fit for small *T*, using genetic algorithm

Table 1 summarises and compares the results obtained from different GA evolutions and the continuous approximation for $N = 10$, i.e. $p_0 = 0.05$, with small *T* values up to 11. For these parameters the predicted total intensity from continuous approximation is 3.80 regardless of trajectory, in the other words, regardless of the objective functions of the GA evolution. When equalising intensities across generations, the precision of predicting total intensity was very good initially with an error of 1.7% at $T = 2$, deteriorating as *T* increases, and then improving again, with the greatest error of predicting the total intensity being 9.2% at $T = 5$. The continuous approximation introduces marginally greater errors to the predicted total intensity when minimising sum of squared intensities, with errors peaking at 12.3% for T = 5 and reducing to 10.4% for T = 11. Note the similar trend on the goodness of fit of the continuous approximation varies with T for both objectives. A very different trend was observed for the objective of minimising total intensity, with total intensity continuing to reduce with T to 3.06 at T = 11 which is very different from the prediction of 3.80. Reasons leading to this observation will be explained in the discussion section.

Looking at the profile of these different GA evolutions reveals more detail about them. The intensity profile of equalising intensities objective is quite similar to minimising sum of squared intensities objective with their intensity achieved each generation became more and more uniform over time (Figure 1a and 1b). This illustrates the derivation showing that the solutions for the two objectives converge given the validity of the continuous approximation.

Assuming the convergence of objectives of equalising intensities and minimising sum of squared intensities, the minimum sum of squared intensities predicted from the continuous approximation is equal to $3.80^2/T$ since $(\frac{i}{T})^2\ T = \frac{i^2}{T}$. Figure 1a shows that as T increases up to 11, the selection intensities become much more uniform, although Table 1 shows the prediction of minimum sum of squared intensities still has significant error at T = 11 despite that the magnitude of the error is reducing. It might be expected that minimising the sum of squared intensities will have approximately twice the error of minimising total intensity (see Appendix 2).

(ii) Goodness of fit for large *T*, using simulations

The simulations allowed the goodness of fit to be tested for large *T* by varying the selection intensity applied. For the results presented in this section, p_0 is 0.001 (*N*=500), with the predicted total intensity being 4.35 from the continuous approximation.

Discrete Generation with Constant Selection Intensity

The comparisons between simulation of breeding with discrete generation and the continuous approximation for a range of different but constant selection intensities applied are shown in Figure 2. The results are presented as the mean number of cohorts required to fixation with the expected number of cohort being calculated by dividing the expected total intensity with the constant intensity applied during the simulation. Figure 2 shows that for constant intensities > 0.5, where T < 10, the scale of errors agrees with the result shown in Table 1. However, the simulations show that the approximation fits the results progressively more closely for all intensities < 0.5. For all intensities < 0.75, the differences between prediction and actual results are less than one cohort.

Discrete Generation with Oscillating Selection Intensity

The independence of the total intensity applied to trajectory was further tested by oscillating selection intensities across cohorts as in a saw tooth pattern. Table 2 shows comparison of total intensity applied for oscillating selection intensities patterns compared to constant selection intensity with same pair-wise average. Results show that the prediction errors are only slightly larger for oscillating selection intensities compared to constant selection intensities with comparable average selection intensity. The approximation still provides good prediction under such conditions, with errors around 2.3% for oscillating selection intensities {0.3, 0.1} and increased to 11.7% for selection intensities {0.6, 0.4}. The increase in error with higher selection intensities and lower fixation times would be expected from the result of constant selection intensities. There were only small differences between complementary patterns, i.e. {0.3, 0.1} compared to {0.1, 0.3} (results not shown).

In all breeding simulations, the prediction often appears as an under-estimation of the simulated result, which is unsurprising because that the selection intensity applied in the simulation could not always be achieved, i.e. in the last few cohorts the target p could exceed 1.0 in order to achieve the selection intensity applied – which is not possible. This is particularly important for large i selection, when only small selection intensity might have been required to move the frequency to 1.

Overlapping Generation

Table 3 summaries the results for simulations with overlapping generations. It shows total intensity required for fixation is predictable from the continuous approximation for low selection intensities but the % errors increase as selection intensity applied per cohort increases. When the unmodified Equation (6) was used, the result is almost identical to those shown in Figure 2. However the use of V_{total}, which represents the full genetic variance, introduces an additional error. The 8.4% error for intensity of 0.5 represents approximately 1 cohort difference between predicted and observed time to fixation. In this case the mean actual number of cohorts was 9.4.

4. DISCUSSION

The theory developed in this paper shows that providing the continuous approximation is valid, then the total intensity applied to move between two frequencies is directly proportional to the difference between the arcsines of (1-2p) for the end points p_0 and p_T irrespective of trajectory – including standard logistic trajectories, $dp/dt = sp(1-p)$. For fixation of a rare mutant, a frequent subject of interest, as p_0 tends to 0 and p_T tends to 1, the total intensity tends to $\sqrt{2}\pi$. Further the strategies of (i) equalising selection intensities throughout the trajectory (Sonesson & Meuwissen, 2004), and (ii) minimising the sum of squared intensities (Sanchez et al, 2006) converge to the same optimal trajectory which is a function of time described by a segment of a sine wave. The results showed that the goodness of fit of the continuous approximation became progressively better as T increased, with prediction errors for total intensity reducing becoming reasonable as $T \sim 10$, or average $i \sim 0.4$ during the period. Further this result remained true for trajectories in which i was varied over time rather than constant, or where generations were overlapping rather than discrete.

The continuous approximation will have a lack of fit for two reasons. First, a smooth curve is used to approximate a step function; second, the dominator for $p_{t+1} - p_t$ in i_t is related to $p_t(1-p_t)$, not $p_{t+\frac{1}{2}}(1-p_{t+\frac{1}{2}})$ which would be more natural for the use of the continuous approximation. This affects the goodness of fit under positive selection since

i_t is greater than expected from approximation by $p'(t+\frac{1}{2})$ when $p(t) < p(t+\frac{1}{2}) < 0.5$, but less than the approximation when $p(t+\frac{1}{2}) > p(t) > 0.5$. The sizes of error are comparable for the pair of $p(t)$ that are in equal deviation from 0.5. These trends are most extreme for p close to 0 or 1, or for small T when $p(t)$ changes rapidly, and there are greater opportunities for cancelling when trajectories move from $p < 0.5$ to $p > 0.5$.

The difference in the sign of errors when p is greater than or less than 0.5 helps to explain the results found for minimising the total intensity, since for all T a trajectory with total intensity less than predicted by the continuous approximation can be found (Table 1). Figure 3 shows the trajectories that minimise total intensity shown in Table 1, and it is seen the trajectory resembles a continuous curve for $p < 0.5$ with a jump in the final generation from close to 0.5 directly to 1. As T increases this represents a discontinuity in the trajectory, which can be seem as a combination of the continuous approximation from p_0 (assumed < 0.5) to 0.5 and a direct jump from 0.5 to 1, For $T = 11$ used in Table 1, the expected value from the discontinuous solution is 3.02 (c.f. 3.06) affirming the continuous approximation can fit well to intervals that do not span both sides of 0.5.

.

The existence of the discontinuous solution for minimising the total intensity creates a distinction between minimising sum of squared intensities and equalising selection intensities. The sum of squared intensities can be broken down into two components: the sum of selection intensity and the variance of selection intensity:

$$\sum_{t=1}^{T} {i_t}^2 = T\ E[i_t]^2 + T\ Var(i_t) = T^{-1}(\sum_{t=1}^{T} i_t)^2\ +\ T\ Var(i_t)$$

The strategy of equalising selection intensities promotes reduction in the sum of squared intensities by having no variance term, while the discontinuous solution is effective through reducing total intensity. For small *T*, the trajectory minimising the sum of squared intensities is temporarily effective in reducing the sum of squared intensities by reducing the total intensity acquired and therefore allowing some variance. However, as *T* increases the benefits from reducing total intensity become less than the penalty from the variance among the selection intensities, and the optimum trajectory moves towards the trajectory of equalising intensities (see Appendix 4).

Genomics is at the start of giving values to many small segments of chromosomes, sometimes with QTL identified, and sometimes simply marked. Simultaneously we are also at the threshold of being able to manage inbreeding at the level of the segment, i.e. requiring slow change in diversity, or wishing to reduce the impact of negative LD on what segments can be fixed in the population. Therefore we envisage the field of “designer genomes” where the target trajectories of multiple loci are mapped out on a genome-wide scale. This is not a problem with only one locus. However, to achieve targets on frequency and inbreeding at multiple loci we need to understand in the long term what is required to fix/eradicate an allele or to move from a frequency point to another, and hence consider how closely the designed genome can be achieved. It is precisely this approximation that allows such predictions over time to be made in a simple fashion albeit that it is but one step towards achieving the wider goal.

One of the possible uses of this approximation is on the removal of the recessive mutant allele that causes foal immunodeficiency syndrome (FIS), more commonly known as the Fell pony syndrome. This fatal condition affects not only Fell ponies but also Dales ponies, and the causal mutant has recently been identified (personal communication: June Swinburne). Although the eradication of this mutant allele is highly desirable, two reasons makes the execution difficult: first, the frequency of carrier is high within the population (~0.4 in the Fell breed, personal communication: June Swinburne), and second, the Fell breed is a small breed. In the other word, this allele is wide-spread in a small gene pool; hence options such as culling of all carriers are not sensible as they might lead to the loss of genetic diversity and the emergence of new recessives. Therefore it is necessary to plan the removal of this mutant allele over a prior time scale to minimise the impact on diversity. Theoretically the process of eradication should be carried out slowly and carefully in order to minimise the reduction on genetic diversity within the breeds. The approximation in this study can provide a simple means of getting a series of stage goals for moving the frequency to zero, i.e. target frequency points, to be achieved over the pre-determined horizon whilst minimising the diversity loss. With the mutant allele frequency ~0.25, the total intensity required to remove the mutant allele is ~1.48, and the intensity in each generation is 1.48/T.

Aspects of the results may be generalised to more than one QTL, and there is a synergy with the results of Goddard (2009), where trajectories for two QTL are optimised with respect to a profit function. The study of Goddard recognises allele frequencies do not change linearly with the selection intensity applied, and uses a transformation to a scale (denoted z in the paper) upon which linearity holds − this requires the continuous approximation to hold since derivatives are required. Appendix 3 shows that the scale, $z(p)$, can be interpreted as being directly proportional to accumulated selection intensity applied to the locus for moving from an infinitesimally small frequency to p. This study shows that to move m loci from $\boldsymbol{p_0}$ to $\boldsymbol{p_T}$ whilst minimising inbreeding at a neutral locus (and one that is affected by selection through the development of the pedigree only) constant selection intensity is required to be simultaneously applied at each locus – albeit with intensity differing among loci. This trajectory is represented by straight lines in an m-dimensional z-space with the relative strength of selection on each locus determining direction, and the line is traversed in T segments of equal length. However the actual inbreeding accumulated will depend on T, the size of the population, and also upon the linkage disequilibrium among the loci being selected.

In conclusion, the continuous approximation shows that: (i) the optimising approaches of equalising intensities (Meuwissen & Sonesson, 2004) and minimising sum of squared intensities (Sanchez et al., 2006) have the same limiting form and converge over time to a sine wave; and (ii) the total intensity required to move an allele from a given frequency point to another can be very closely approximated and only depends on the starting and end frequency.

Acknowledgements

The BBSRC are gratefully acknowledged for funding.

References

Dekkers, J. C. M., and R. Chakraborty, 2001 Potential gain from optimizing multigeneration selection on an identified quantitative trait locus. *Journal of Animal Science* **79**: 2975-2990.

Dekkers, J. C. M., and J. A. M. van Arendonk, 1998 Optimizing selection for quantitative traits with information on an identified locus in outbred populations. *Genetical Research* **71**: 257-275.

Gibson J.P. Short term gain at the expense of long term response with selection on identified loci. Proceedings of the Fifth World Congress on Genetics Applied to Livestock Production , 201-204. 1994. 7-8-1994.
Ref Type: Conference Proceeding

Meuwissen, T. H. E., and A. K. Sonesson, 2004 Genotype-assisted optimum contribution selection to maximize selection response over a specified time period. *Genetical Research* **84**: 109-116.

Pong-Wong, R., and J. A. Woolliams, 1998 Response to mass selection when an identified major gene is segregating. *Genetics Selection Evolution* **30**: 313-337.

Sanchez, L., A. Caballero, and E. Santiago, 2006 Palliating the impact of fixation of a major gene on the genetic variation of artificially selected polygenes. *Genetical Research* **88**: 105-118.

Shepherd, R. K., and B. P. Kinghorn, 1992 Optimizing Multitier Open Nucleus Breeding Schemes. *Theoretical and Applied Genetics* **85**: 372-378.

Villanueva B., Dekkers J.C.M., Woolliams J.A. & Settar P. Maximising genetic gain with QTL information and control of inbreeding. Proceedings of the Seventh World Congress on Genetics Applied to Livestock Production . 2002. 19-8-2002.
Ref Type: Conference Proceeding

Villanueva, B., J. C. M. Dekkers, J. A. Woolliams, and P. Settar, 2004 Maximizing genetic gain over multiple generations with quantitative trait locus selection and control of inbreeding. *Journal of Animal Science* **82**: 1305-1314.

Weisstein E.W. Euler-Lagrange differential equation. http://mathworld.wolfram.com/euler-lagrangedifferentialequation.html . 2005. 4-10-2006.
Ref Type: Electronic Citation

Woolliams, J. A., and P. Bijma, 2000 Predicting rates of inbreeding: in populations undergoing selection. *Genetics* **154**: 1851-1864.

Woolliams, J. A., N. R. Wray, and R. Thompson, 1993 Prediction of Long-Term Contributions and Inbreeding in Populations Undergoing Mass Selection. *Genetical Research* **62**: 231-242.

Table 1. The total intensity (*Sum* i_t) and sum of squared intensities (*Sum* i_t^2) required for N=10 and a range of T values, using three optimisation strategies: 1) equalising selection intensities across generations, 2) minimising *Sum* i_t^2, and 3) minimising *Sum* i_t, and 4) calculated from the continuous approximation.

Strategy	**Criterion**	***T*=2**	**% error**	***T*=5**	**% error**	***T*=8**	**% error**	***T*=11**	**% error**
Equalise i_t	*Sum* i_t	3.869	-1.7	3.456	9.2	3.474	8.7	3.509	7.8
	Sum i_t^2	7.486	-3.4	2.390	17.5	1.509	16.6	1.119	15.0
Minimise *Sum* i_t^2	*Sum* i_t	3.790	0.4	3.338	12.3	3.364	11.6	3.411	10.4
	Sum i_t^2	7.296	-0.8	2.300	20.6	1.460	19.3	1.088	17.3
Minimise *Sum* i_t	*Sum* i_t	3.748	1.5	3.166	16.8	3.088	18.8	3.059	19.6
	Sum i_t^2	7.670	-6.0	3.154	-8.9	2.604	-43.9	2.404	-82.7
Prediction	*Sum* i_t	3.805		3.805		3.805		3.805	
	Sum i_t^2	7.239		2.896		1.810		1.316	

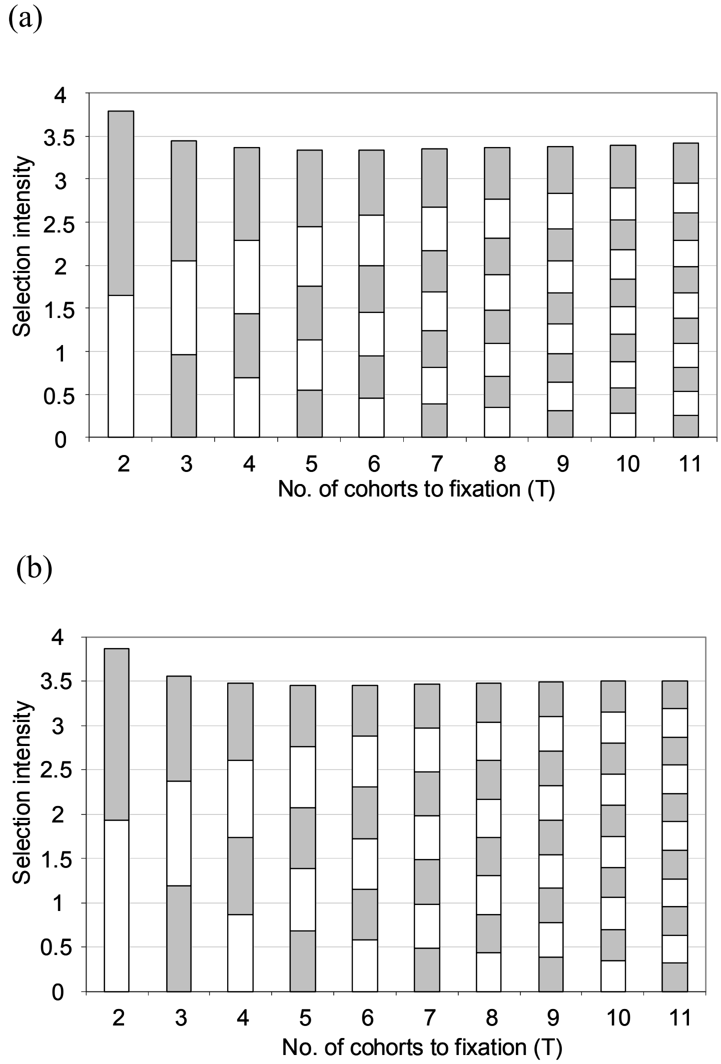
Table 2. Comparison between the total intensity (*Sum* i_t) required to fix an allele under simulations with discrete generations and predicted from continuous approximation for a range of different selection intensities. The selection intensity can be either constant all through the simulation or oscillating between a pair of different values (shown as {a, b}). Population size (N) equals 500 in all cases.

Selection Intensity	**0.2**	**{0.3,0.1}**	**0.3**	**{0.4,0.2}**	**0.5**	**{0.6,0.4}**
Predicted *Sum* i_t	4.35	4.35	4.35	4.35	4.35	4.35
Simulated *Sum* i_t	4.40	4.45	4.50	4.55	4.68	4.86
% error	1.1	2.3	3.4	4.6	7.6	11.7

Table 3. A comparison between the total intensity (*Sum* i_t) required to fix an allele in simulations with overlapping generations for different constant selection intensities applied and for genetic variance calculated by different methods. In "Unmodified" Equation (6) was used directly, but in "Modified" the true genetic variance V_{total} replaced $\frac{1}{2}p_t(1-p_t)$ in Equation (6). In all cases population size (*N*) equals 500 and predicted *Sum* i_t = 4.35. Standard errors, % error in prediction, and generation interval (*L*) are also shown.

	Selection intensity/cohort = 0.2		**Selection intensity/cohort = 0.5**	
	Unmodified	*Modified*	*Unmodified*	*Modified*
Sum i_t	4.33 ± 0.012	4.30 ± 0.006	4.51 ± 0.016	4.72 ± 0.012
% error	-0.5	-1.1	3.7	8.4
L	2.29	2.32	2.06	2.21

Figure 1. The composition of total intensity obtained from GA with the objective of (a) minimising sum of squared intensities and (b) equalising selection intensities. Each block represents the amount of selection intensity achieved in a single mating/frequency change. Shading is for the purpose of illustration only.



(a)

(b)

Figure 2. Comparison between the numbers of cohorts required to fix an allele for a range of different selection intensities, for (a) simulations with discrete generation (open circle) and (b) continuous approximation (filled circle). Population size (*N*) equals 500 in all cases. The standard deviations are shown as error bars and the standard errors are negligible.

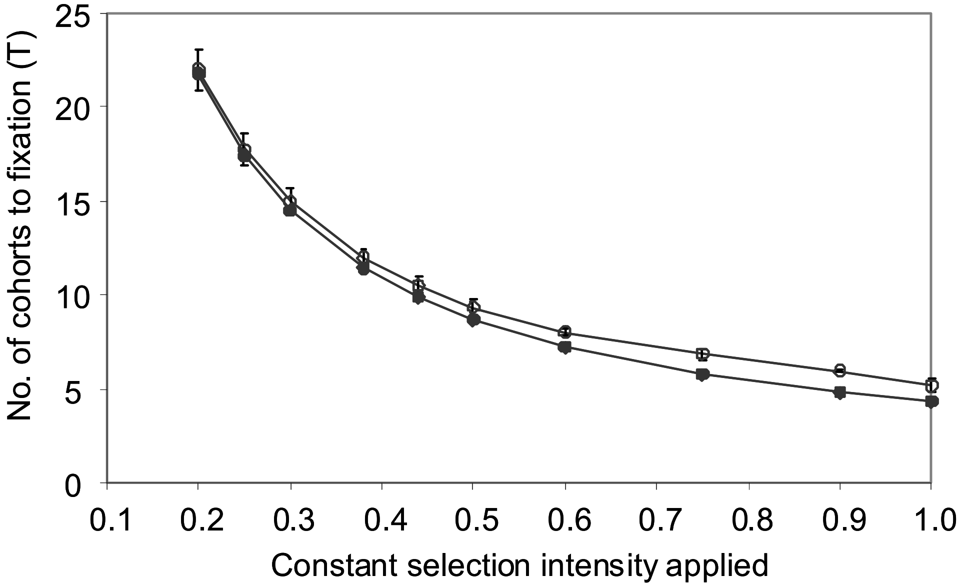
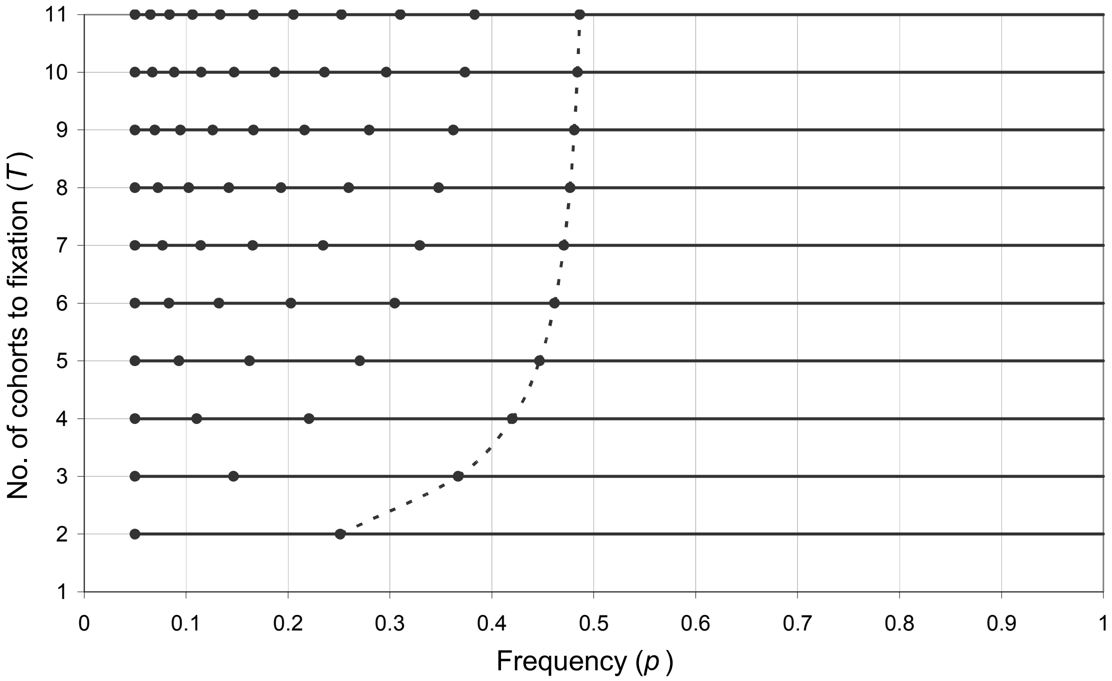


Figure 3. The frequency path (trajectory) obtained by GA with the objective of minimising total intensity for different *T* values. For each profile with different *T*, the frequency points are shown as the solid circles along the horizontal line, with the first frequency point being at *p*=0.05. The last frequency points, p_{T-1}, from all profiles are joined by dashed line to illustrate how p_{T-1} approaches 0.5 as *T* increases.



Appendix 1

Minimising the Total Intensity

By noting that $p'(t)dt$ can be replaced by dp, and then substituting *p* for *p(t)*, Equation (2) can be transformed to the following equation, where its direct integration leads to Equation (3):

$$\sum_{0}^{T} i_t \approx \int_{p_0}^{p_T} \frac{dp}{\sqrt{\frac{1}{2}p(1-p)}} \tag{A1}$$

Minimising Properties of Allele Trajectories

An important methodology for optimising trajectories is the calculus of variations (Weisstein 2005). When the function to be optimised is of the form:

$$\int_{0}^{T} f[p, p', t]dt \tag{A2}$$

then the solution can be obtained from the Euler-Lagrange equations providing trajectories *p(t)* are differentiable. This equation states that the optimum trajectory satisfies:

$$\partial f/\partial p - d[\partial f/\partial p']/dt = 0$$

This solution can be further simplified if $f[p, p', t]$ is independent of explicit dependence on *t*, i.e. the partial derivative of f[ ] with respect to *t* is 0 (i.e. $\delta f/\delta t = 0$), then the condition may be simplified to the Beltrami identity:$f[p, p', t] - p'\delta f/\delta p' = C$, where *C* is a constant of integration.

Minimising Sum of Squared Intensity: The function f[ ] required to minimise the sum of squared intensity is as follows:

$$\int_0^T f[p,p',t]dt = \int_0^T \frac{p'^2}{0.5p(1-p)}dt \tag{A3}$$

where *p'* is the derivative of *p* with respect to *t*, and $f[p,p',t] = p'^2[0.5p(1-p)]^{-1}$, representing the square of the selection intensity at time *t*. Applying the method of calculus of variation (Weisstein 2005) to the sum of squared intensities gives the following result: $f[p,p',t] - p'\delta f/\delta p' = -p'^2[0.5p(1-p)]^{-1} = C$ and *p* must satisfy:

$$p' = [0.5Cp(1-p)]^{\frac{1}{2}} \tag{A4}$$

Solving this differential equation gives $\sin^{-1}(1-2p) = At + B$ or, equivalently, $p(t) = \frac{1}{2}[1-\sin(At+B)]$, where *A* and *B* are constants of integration. This comes from noting that $[p(1-p)]^{\frac{1}{2}} = \frac{1}{2}[1-u^2]^{\frac{1}{2}}$, where $u = (1-2p)$ to convert the function into a recognizable standard integral form. *A* and *B* are determined by the desired change from $t = 0,\cdots,T$ and have units of radians (not degrees). The optimal trajectory for minimising the sum of squared intensity applied to the allele is therefore a segment of a sine wave.

Equalising Selection Intensities: The objective function of equalising the selection intensities is equivalent to making the selection intensity constant i.e. $p'[0.5p(1-p)]^{-\frac{1}{2}} = C$, which is the same differential equation as that obtained above for the criterion of minimising sum of squared intensities (Equation A4).

Appendix 2

Because the error associated with minimising sum of squared intensity (as a percentage to the total), can be simplified as: $\frac{(i_t+\delta(i_t))^2-i_t{}^2}{i_t{}^2} \approx \frac{2\imath_t\delta(i_t)}{i_t{}^2} = \frac{2\delta(i_t)}{i_t}$ given $(\delta(i_t))^2$ can be neglected; which is twice the error of minimising total intensity ($\frac{\delta(i_t)}{i_t}$).

Appendix 3

This study considers the total intensity required to move from p_0 to p_T, as:

$$\int_{p_0}^{p_T} \frac{dp}{\sqrt{0.5p(1-p)}}$$

Note that in Goddard (2009) z(p) $= \sin^{-1}(\sqrt{p}) = \pi/4 - \frac{1}{2}\ sin^{-1}(1-2p)$ and that

$$\int_{p_0}^{p_T} \frac{dp}{\sqrt{0.5p(1-p)}} = \int_{z_0}^{z_T} \frac{dp}{dz}\frac{dz}{\sqrt{0.5p(1-p)}} = \int_{z_0}^{z_T} dz = z_T - z_0$$

Therefore the increment in z is the accumulated selection intensity applied to the locus.

Appendix 4

For large *N*, the total intensity for the continuous solution$\approx \sqrt{2}\pi$, while the total intensity for the discontinuous solution approaches $\sqrt{2}(1+\pi/2)$. This is obtained by calculating separately the intensity from p_0 (assumed < 0.5) to 0.5 and the intensity from 0.5 to 1. The first of these is $\sqrt{2}\ sin^{-1}(1-N^{-1})$, which tends to $\pi/\sqrt{2}$ as N becomes large, whilst the second is $\sqrt{2}$, giving a minimum of $\sqrt{2}(1+\pi/2)$ for large *N*.

Hence, for continuous solution the sum of squared intensities $\sum i^2 = T(\frac{\sqrt{2}\pi}{T})^2 = \frac{2\pi^2}{T}$ and for discontinuous solution $\sum i^2 = (\sqrt{2})^2 + (T-1)(\frac{\frac{\pi}{\sqrt{2}}}{T-1})^2 = 2 + \frac{\pi^2}{2(T-1)}$.

The sum of squared intensities from the two solutions for a range of *T* values are summarised below. When *T* = 7, the two solutions yield roughly equal results, and for *T* > 7, the continuous solution performs better than the discontinuous solution.

e	**Continuous**	**Discontinuous**
2	9.87	6.93
3	6.58	4.47
4	4.93	3.64
5	3.95	3.23
6	3.29	2.99
7	2.82	2.82
8	2.47	2.70
9	2.19	2.62
10	1.97	2.55