Theoretical Population Biology

How single-locus FST varies with migration and selection for two populations in migration-selection balance
--Manuscript Draft--

Manuscript Number:	TPB-D-21-00048
Article Type:	Full Length Article
Section/Category:	Genetics
Keywords:	migration-selection balance; dispersal; population genetics; ecological genomics; FST
Abstract:	A complete understanding of the role of natural selection in driving evolutionary change requires accurate estimates of the strength of selection acting at the genetic level in the wild. This is challenging to achieve but may be easier in the case of populations in migration-selection balance. When two populations are at equilibrium under migration-selection balance, there exist loci whose alleles are selected different ways in the two populations. Such loci can be identified from genome sequencing by their high values of F_{ST}. This raises the question of what is the strength of selection on locally-adaptive alleles. To answer this question we analyse a deterministic 1-locus 2-allele model of a large population distributed between two niches. We derive theory showing the relationships between values of F_{ST}, equilibrium allele frequencies, selection coefficients, migration rates and dominance. We illustrate our results with graphs showing how, at a given locus, F_{ST} increases with the selection coefficients acting on the alleles at the locus, for given migration rates and dominance relationships between the alleles. Given the extent of recent progress in ecological genomics, we hope our methods may help those studying migration-selection balance to quantify the advantages conferred by adaptive genes.

How single-locus F_{ST} varies with migration and selection for two populations in migration-selection balance

Abstract

A complete understanding of the role of natural selection in driving evolutionary change requires accurate estimates of the strength of selection acting at the genetic level in the wild. This is challenging to achieve but may be easier in the case of populations in migration-selection balance. When two populations are at equilibrium under migration-selection balance, there exist loci whose alleles are selected different ways in the two populations. Such loci can be identified from genome sequencing by their high values of F_{ST}. This raises the question of what is the strength of selection on locally-adaptive alleles. To answer this question we analyse a deterministic 1-locus 2-allele model of a large population distributed between two niches. We derive theory showing the relationships between values of F_{ST}, equilibrium allele frequencies, selection coefficients, migration rates and dominance. We illustrate our results with graphs showing how, at a given locus, F_{ST} increases with the selection coefficients acting on the alleles at the locus, for given migration rates and dominance relationships between the alleles. Given the extent of recent progress in ecological genomics, we hope our methods may help those studying migration-selection balance to quantify the advantages conferred by adaptive genes.

Keywords: migration-selection balance; dispersal; population genetics; ecological genomics; F_{ST};

Introduction

A complete understanding of the role of natural selection in driving evolutionary change requires accurate estimates of the strength of selection acting at the genetic level in the wild. Until recent advances in molecular population genetics, measuring natural selection at the genetic level has been challenging (Linnen and Hoekstra, 2009; Thurman and Barrett, 2016). (Thurman and Barrett, 2016) located 79 papers that used molecular techniques to study natural selection acting at the genetic level in natural populations, from which the importance of genomic data is clear, though variation in time and space can complicate tracking the strength of selection (Rudman et al., 2018). Several methods for inferring the strength of selection from gene frequency data have been developed in recent years (Tataru et al., 2017). Starting with the Wright-Fisher model of the effects of random genetic drift in a randomly mating population of finite size, several approaches have used the diffusion approximation to estimate the effects of various combinations of mutation, migration and selection on how allele frequencies change over the generations. For example (Vitalis et al., 2014) introduced a method extending the diffusion approximation of genetic drift in the migration-drift equilibrium island model to allow for the effects of selection. When applied to analysis of selection on the lactase-producing gene LCT (Vitalis et al., 2014)'s method showed that the strongest selection coefficients occurred in Europe and the Indus Valley, where scaled selection coefficients ranged up to 100. Scaled selection coefficients are defined as the products of selection coefficient and deme population size, but their effects can generally not be separated: to estimate selection coefficients, further information is needed on deme population sizes. Allele frequencies under the Wright–Fisher model can be approximated as time goes to infinity, but this requires that the scaled parameters relating to migration, mutation and selection are on the order of 1 (Tataru et al., 2017).

These studies started with a model of genetic drift and required complex mathematical development. Some simplification can be achieved if the starting point is instead a large population, so that drift can be ignored. The advantages of this approach have been investigated by (Jewett et al., 2016), who conclude that "ignoring drift leads to estimates of selection coefficients that are nearly as accurate as estimates that account for the true population history, even when population sizes are small and drift is high. This result is of interest because inference methods that ignore drift are widely used in evolutionary studies and can be many orders of magnitude faster than methods that account for population sizes." (Hoekstra et al., 2004) followed this approach of ignoring drift. Starting from models of populations in migration-selection balance (Haldane, 1930; Wright, 1931), they derived for a simple 2-allele 2-population model an equation that allows estimation of the selection coefficient against a deleterious allele from its equilibrium frequency in each of two populations, the level of dominance and the migration rates between the populations. Using the equation, (Hoekstra et al., 2004) estimated the selection coefficient acting on the *Mc1r* gene, which codes for coat colour, in populations of pocket mice living on black lava and on neighbouring light rocks. Here we use a similar approach to study locally adapted populations at equilibrium, held in a balance with selection acting in different directions in different populations. This results in genetic differentiation between the populations if migration rates between the populations are sufficiently low. We show how, if drift can be ignored, genomic studies of local adaptation can be used to estimate selection coefficients acting on the individual genes which together confer local adaptation.

Genetic differentiation between populations is generally summarised by F_{ST} (Whitlock, 2011), and (Beaumont and Nichols, 1996) showed how F_{ST} can be used in genomic studies to identify loci responsible for local adaptations by their relatively high values of F_{ST} (for recent examples see, e.g., (Flanagan and Jones, 2017; Graham et al., 2018; Savolainen et al., 2013)). Loci not under selection have relatively low values of F_{ST} and information from these loci can provide estimates of migration rates between niches. As an example consider (Graham et al., 2018)'s study of high and low altitude populations of the speckled teal (*Anas flavirostris*) in South America, which used genome sequencing and F_{ST} analysis to identify 'outlier' genes selected in opposite ways in the two populations. The outlier genes had F_{ST} values in the range 0.44 – 0.77, and in some cases it was possible to identify the

functions of the outlier genes that are adaptive at high altitude. The remaining non-outlier genes had F_{ST} values around 0.05 which, when further analysed, suggested < 3% migration rate from low to high altitude, and less from high to low. This genomic study of local adaptation identified loci responsible for local adaptations by their high values of F_{ST}, and genomic variation in unselected regions of the genome to estimate migration rates between niches. High values of F_{ST} occur in genes responsible for local adaptation in large populations at equilibrium, but F_{ST} may also be high in populations not at equilibrium, in which evolution is still under way, in ways that depend on the spatial structure of the environment (e.g., island model, isolation-by-distance, range expansion)(Lotterhos and Whitlock, 2014; Lotterhos and Whitlock, 2015). The methods developed in the present paper assume that populations are at equilibrium.

F_{ST} is calculated from measurements of the frequency of alleles of each type that are present in each population. Intuitively it is clear that F_{ST} is increased by the strength of selection in each population but eroded by migration between them, but a method is needed to quantify the relationship. As a first step, we show here how (Hoekstra et al., 2004)'s equation for a deterministic 1-locus 2-allele model of a large population distributed between two niches can be corrected and extended to show how selection coefficients can be obtained for both niches. The population is in migration-selection balance. Account is taken of migration rates and levels of dominance, and mating takes place within niches as in (Bulmer, 1972). Our derivation assumes equal population sizes in the two niches, but is easily modified for the general case. We show graphically how F_{ST} increases with selection coefficients for given migration rates and dominance relationships between the alleles. Equilibrium is only possible for restricted values of selection coefficients. For a given value of F_{ST}, the possible values of the selection coefficients, designated s_1 and s_2, lie on a contour on the $F_{ST}(s_1, s_2)$ surface. F_{ST} is here calculated from equilibrium allele frequencies, and we also provide equations for the general relationships between equilibrium allele frequencies, selection coefficients, migration rates and dominance. Without these equations, equilibrium frequencies can only be ascertained by simulation of the evolutionary process, as in (Sibly and Curnow, submitted). Given current interest in studying local adaptation using ecological genomics, we hope our methods may help quantify the advantages conferred by adaptive genes.

Theory and methods

Population genetic model

Our model has one locus with two alleles P and Q in an environment consisting of two niches with some migration between niches prior to mating, as depicted in Fig. 1. The locus determines ecological adaptation to one niche or the other. The fitness of the three genotypes in each niche is shown in Table 1.

	Niche 1			Niche 2		
Genotype	PP	PQ	QQ	PP	PQ	QQ
Fitness	1	$1+hs_1$	$1+s_1$	1	$1+hs_2$	$1+s_2$
Relative frequency	p_1^2	$2p_1q_1$	q_1^2	p_2^2	$2p_2q_2$	q_2^2

Table 1. The fitness of the three genotypes in each niche. p_1 and q_1 represent the relative frequencies of the P and Q alleles at the start of a generation in niche 1, their frequencies in niche 2 are p_2 and q_2. Carriers of the QQ genotype obtain fitnesses $1 + s_1$ and $1 + s_2$ in niches 1 and 2 respectively, PP homozygotes have fitness 1 in both niches. Parameter h indicates the level of dominance of the Q allele.

We assume a large population so that the dynamics are deterministic. Generations are discrete and individuals die after mating. The life histories occur in the following order. At the start of each generation individuals in each niche mate at random, and all mating individuals obtain the same number of offspring. The number of offspring of each genotype that survive in each niche is the product of its initial frequency and its fitness. Population regulation then returns population numbers to their initial values so that both niches have the same number of individuals. Finally some individuals migrate between niches, as shown in Fig. 1, which leads to the start of the next generation.
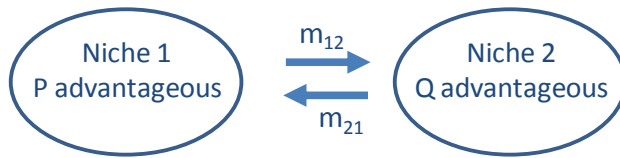

Fig. 1. Conceptual overview of the model. For clarity the equal-size niches are shown distinct, but in nature may be contiguous or overlap. m_{12} and m_{21} specify the proportion of individuals in one niche that migrate to the other each generation after viability selection and population regulation have occurred. When analysing the model we suppose that Q is disadvantageous in niche 1 (i.e., s_1 is negative) but advantageous in niche 2 (i.e., s_2 is positive), while PP homozygotes have fitness 1 in both niches. There are no sex differences in fitnesses or migration rates.

Recurrence equations giving the frequencies of the P allele in successive generations

We now derive recurrence equations for the above model that give the frequencies of the P allele in successive generations in each niche. The relative frequencies of P and Q at the start of a generation in the two niches are given in Table 1. The relative frequencies of P in the two niches after selection are:

$$p_1' = (p_1^2+(1+hs_1)p_1q_1)/(1+2hp_1q_1s_1+s_1q_1^2) \tag{1a}$$

$p_2' = (p_2{}^2+(1+hs_2)p_2q_2)/(1+2hp_2q_2s_2+s_2q_2{}^2)$ (1b)

The relative frequencies of P after migration are:

$$p_1'' = \frac{(1-m_{12})p_1'+m_{21}p_2'}{1+m_{21}-m_{12}} = ((1-m_{12})p_1' + m_{21}p_2')/c_1$$ (2a)

$$p_2'' = \frac{(1-m_{21})p_2'+m_{12}p_1'}{1+m_{12}-m_{21}} = ((1-m_{21})p_2' + m_{12}p_1')/c_2$$ (2b)

where p_1' and p_2' are obtained from equations 1, $c_1 = 1 + m_{21} - m_{12}$ and $c_2 = 1 + m_{12} - m_{21}$

The equations show how the frequencies of P in the two niches in the next generation, p_1'' and p_2'', can be derived from the frequencies in the present generation, p_1 and p_2.

Calculation of selection coefficients from equilibrium values of p_1 and p_2 and migration rates

The recurrence equations derived in the previous section can be used to analyse what happens when the population reaches equilibrium. At equilibrium $p_1''=p_1$ and $p_2''=p_2$, so from equations 2:

$$p_1' = \frac{c_1p_1-m_{21}p_2'}{1-m_{12}} = \frac{c_2p_2-(1-m_{21})p_2'}{m_{12}}$$ (3)

If m_{12}, m_{21} and equilibrium values of p_1 and p_2 are known, then equation 3 can be rearranged to obtain:

$$p_2' = \frac{m_{12}(c_1p_1+c_2p_2)-c_2p_2}{m_{12}+m_{21}-1}, = K_2, \text{ say.}$$ (4a)

Similarly $p_1' = \frac{m_{21}(c_1p_1+c_2p_2)-c_1p_1}{m_{12}+m_{21}-1}, = K_1, \text{ say.}$ (4b)

Setting this equal to p_2' as given by equation 1b we obtain:

$$s_2 = \frac{p_2-K_2}{K_2q_2{}^2+hp_2q_2(2K_2-1)}$$ (5a)

$$s_1 = \frac{p_1-K_1}{K_1q_1{}^2+hp_1q_1(2K_1-1)}$$ (5b)

where K_1 and K_2 are given by equation 4. When m_1, m_2 and h are known, uncertainty in the estimates of s_1 and s_2 can be quantified because the estimates of p_1 and p_2 are binomially distributed. Equations 5 show how selection coefficients s_1 and s_2 can be estimated from measurements of m_{12}, m_{21}, h and equilibrium values of p_1 and p_2.

Calculating equilibrium frequencies of alleles in each population when selection coefficients are known

We now turn to the inverse problem of calculating equilibrium frequencies of alleles in each population when selection coefficients are known. From equation 3:

$$p_1 = \frac{m_{12}+m_{21}-1}{c_1m_{12}}p_2' + \frac{1-m_{12}}{c_1m_{12}}c_2p_2$$ (6)

Substituting for p_2' from equation 1b gives an equation for p_1 in terms of p_2, m_{12}, m_{21}, h, s_1 and s_2:

$$p_1 = \frac{m_{12}+m_{21}-1}{c_1m_{12}}\left(\frac{p_2(1+hs_2q_2)}{1+2hs_2p_2q_2+s_2q_2{}^2}\right) + \frac{1-m_{12}}{c_1m_{12}}c_2p_2$$ (7a)

The analogous equation for p_2 is:

$$p_2 = \frac{m_{12}+m_{21}-1}{c_2m_{21}}\left(\frac{p_1(1+hs_1q_1)}{1+2hs_1p_1q_1+s_1q_1{}^2}\right) + \frac{1-m_{21}}{c_2m_{21}}c_1p_1$$ (7b)

Substituting p_2 from equation 7b into equation 7a yields an equation in p_1, m_{12}, m_{21}, h, s_1 and s_2 which can be solved to obtain equilibrium values of p_1 for given values of m_{12}, m_{21}, h, s_1 and s_2. Equilibrium values of p_2 can be obtained similarly. Equations 7 show how equilibrium frequencies of alleles in each population can be calculated when selection coefficients are known.

Finding evolutionary outcomes by simulation of the evolutionary process

Equations 7 do not provide explicit expressions for the equilibrium values of p_1 and p_2. An alternative to using an equation solver to obtain equilibrium values is to simulate the evolutionary process for specified values of h, m_{12}, m_{21}, s_1 and s_2 using the recurrence equations 1 and 2. These relate allele frequencies in one generation to their frequencies in the preceding generation. The simulations whose equilibrium values are presented here were started at an initial frequency of Q of 0.01 in niche 2 and continued to equilibrium as determined by visual inspection of trajectories of allele frequencies (example in Fig. 2). Equilibrium was generally reached within a few hundred generations but took over 10, 000 generations for (s_1, s_2) = (-0.01, 0.01), (-0.01, 0.1) and (-0.1, 0.01) when m = 0.01. Checks showed that the simulated evolutionary outcomes satisfied equations 7.
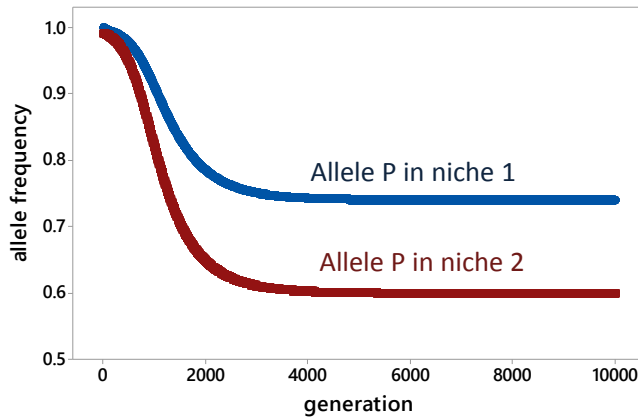

Fig. 2. Example of the simulated evolutionary process for the case h = 1, m = 0.05, s_1 = -0.01, s_2 = 0.01. In this case equilibrium was reached in 6000 generations.

Calculation of F_{ST}

F_{ST} was calculated from allele frequencies using the equation $F_{ST} = \sigma_S^2/\sigma_T^2$, where σ_S^2 and σ_T^2 represent the variances of an allele's frequency between subpopulations, and in the total population, respectively (Holsinger and Weir, 2009). The variance of p, the frequency of an allele in a population of size n, is given by the binomial distribution as npq. So $F_{ST} = (p_T q_T - \frac{1}{2}p_1 q_1 - \frac{1}{2}p_2 q_2)/p_T q_T$, where subscripts T, 1 and 2 refer to the total population and the populations in niches 1 and 2 respectively. F_{ST} can therefore alternatively be thought of in terms of the average frequency of heterozygotes in the two populations compared with the frequency of heterozygotes if there was random mating between all the individuals in the two populations. F_{ST} was assigned the value 0 when one allele became fixed in both niches because F_{ST} tends to 0 as any one allele tends to fixation.

Visualisation of the function $F_{ST} = F_{ST}(m, s_1, s_2)$ when $m_{12} = m_{21} = m$

In this section we show how equilibrium values of F_{ST} are related to the migration rates between the niches, and the strengths of selection s_1 and s_2 within them. We assume from now on that migration rates are the same in each direction, so that $m_{12} = m_{21} = m$. We begin by showing how genetic differentiation, measured by F_{ST}, is eroded by migration when selection is maximal, total selection one way in niche 1 and the other way in niche 2 (Fig. 3). Maximal selection produces maximum

values of F_{ST}, and these decline as migration rates increase, from a maximum of 1 when populations are isolated, to around 0.4 when migration rates are 10% and down to zero when migration rate is 50% (Fig. 3). The maximum values of F_{ST} are a little higher for h=0.5 than for h=1 (dominance) when migration rates are less than 15%, but otherwise similar.
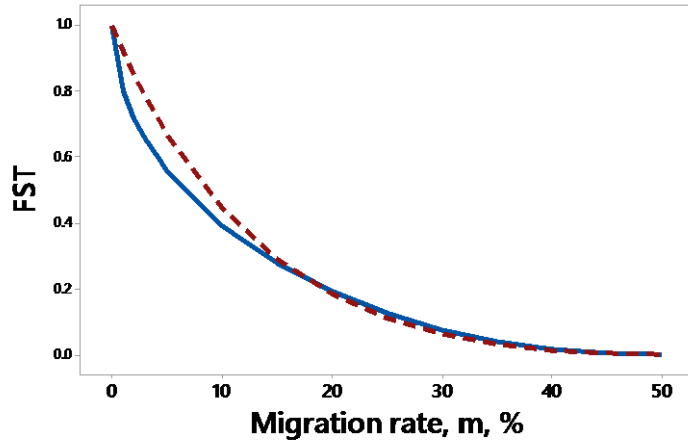

Fig. 3. The maximum values of F_{ST} in relation to migration rate, m. Maximum values of F_{ST} occur when selection on QQ is total one way in niche 1 and the other way in niche 2, and this was approximated by setting s_1 at -1 and s_2 at 10^9. Solid blue curve is for h=1, dashed red curve h = 0.5. Values of F_{ST} were calculated for populations at equilibrium as determined by simulation of evolutionary trajectories of allele frequencies. Equilibrium judged by eye was generally achieved within 100 generations. The table of F_{ST} vs. m from which the figure was constructed is given in Supplementary Materials.

To visualise the effects of lower values of s_1 and s_2 on F_{ST} we calculated equilibrium values of F_{ST} over a grid of values of m, s_1 and s_2 to obtain visualisations of the function $F_{ST} = F_{ST}(m, s_1, s_2)$, and these are presented in Fig. 4 for two levels of dominance h. The panels in the rows of Fig. 4 represent visualisations of the function $F_{ST} = F_{ST}(m, s_1, s_2)$ for three values of migration rate m. The highest values of F_{ST} are achieved when migration rates are lowest (left-hand panels of Fig. 4). Within each panel the highest value of F_{ST} occurs when selection is at its strongest, for Q in niche 2 ($s_2 = 1$) and against Q in niche 1 ($s_1 = -1$). If selection is too low, then either P or Q go to fixation and F_{ST} goes to zero. At other values of s_1 and s_2 the equilibrium is a polymorphism and $F_{ST} > 0$. The shapes of the F_{ST} surfaces for h = 1 (top row of Fig. 4) and h = 0.5 (bottom row) are qualitatively similar but they differ a little in quantitative detail. The F_{ST} surfaces shown in Fig. 4 are not perfectly symmetrical about the $s_2 = - s_1$ plane, because $s_1 = -1$ represents total selection against QQ in niche 1 but $s_2 = +1$ does not represent total selection for QQ in niche 2; that is achieved when $s_2 = \infty$.
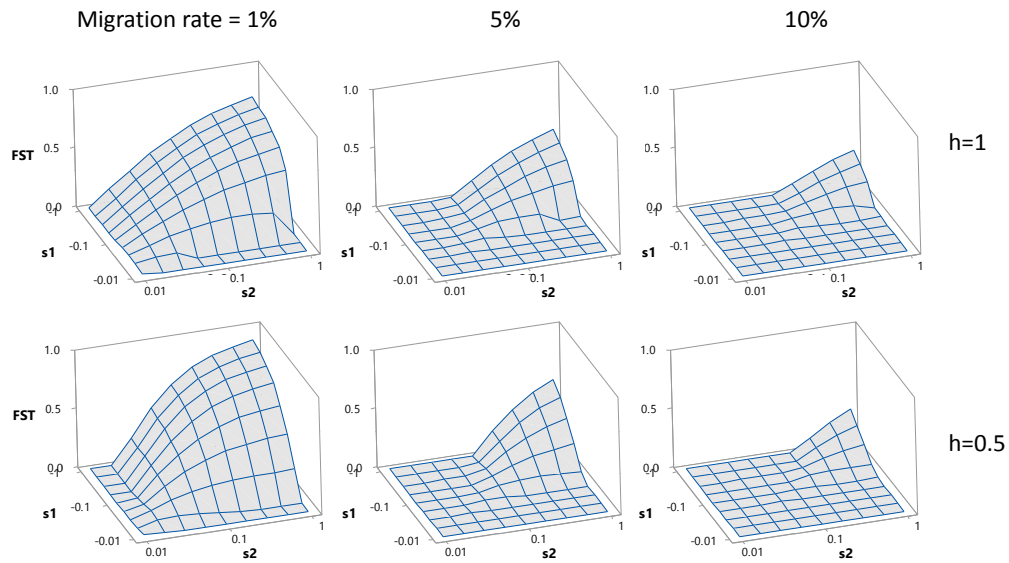
Fig. 4 Values of F_{ST} in relation to selection coefficients s_1 and s_2 for three rates of migration between niches m, and two levels of dominance h. Top row: Q is dominant (i.e. h = 1); bottom row h = 0.5. Values of F_{ST} were calculated for populations at equilibrium as determined by simulation of the evolutionary process. The tables of s_1, s_2 and F_{ST} from which these figures were constructed are given in Supplementary Materials.

If F_{ST} and migration rates are known, Fig. 4 can be used to put bounds on selection coefficients. For a given value of migration rate, possible values of s_1 and s_2 for a given value of F_{ST} lie on the relevant contour in Fig. 4. Contours are shown in Fig. 5 for the case that migration rate = 1% and h=0.5. From the contour corresponding to a given value of F_{ST}, the possible determining values of s_1 and s_2 can now be identified. If $F_{ST} = 0.7$ and selection pressures are equal, for instance, then s_2 and $|s_1|$ must be around 0.25, but if they are not equal then the value of one or other of them could be as low as 0.15.
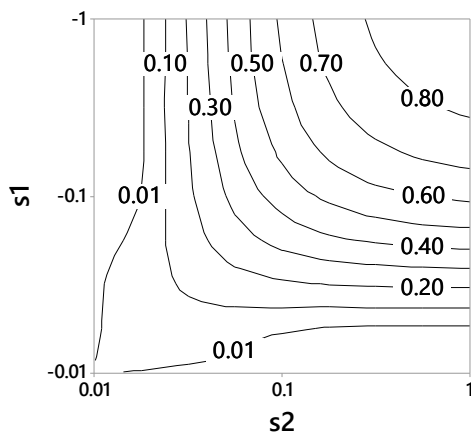

Fig. 5. Contours of F_{ST} in relation to selection coefficients s_1 and s_2 for the case that migration rate = 1% and h=0.5. These are the contours for the lower left panel in Fig. 4.

Discussion

Methods to estimate selection coefficients are needed to further understand evolutionary processes in wild populations. Measurements of F_{ST}, migration rates and allele frequencies can be made relatively easily in locally adapted populations, and selection coefficients can then be estimated if the populations can be assumed to be in equilibrium. Here, using a two-niche two-allele model, we show in equations 5 how selection coefficients s_1 and s_2 can be estimated from measurements of migration rates m_{12} and m_{21}, levels of dominance h and the equilibrium values of allele frequencies in the two habitats, p_1 and p_2. Because F_{ST} is generally reported rather than allele frequencies, we analysed the relationship between F_{ST} and selection coefficients, and present the results graphically in Figs. 3 – 5 as visualisations of the function $F_{ST} = F_{ST}(m, s_1, s_2)$ when migration rates m between the two niches are equal. These visualisations show how equilibrium values of F_{ST} are related to the migration rates between the niches, and the strengths of selection s_1 and s_2 within them. F_{ST} declines as migration rates increase (Figs. 3 and 4). If F_{ST} and migration rates are known, Figs. 4 and 5 show how bounds can be put on possible values of selection coefficients. The analysis presented assumes that after population regulation the populations have equal size. If instead the sizes of populations 1 and 2 are M_1 and M_2 respectively then equations 2 have to be modified with m_{21} replaced by $m_{21}M_2/M_1$ in equation 2a, and m_{12} replaced by $m_{12}M_1/M_2$ in equation 2b, with concomitant modifications in subsequent equations.

Our model agrees with that of (Hoekstra et al., 2004) in modelling selection and corrects an error in (Hoekstra et al., 2004)'s model of migration. Selection acts on migrants as well as residents, but this is not accounted for in (Hoekstra et al., 2004)'s equation 1.

Inferences in practice will need to take account of several caveats. Our calculations are for large populations at equilibrium, but in the real world selection pressures, migration rates and population sizes may vary over time, so that F_{ST} and other variables vary too. The methods presented here need extension to take account of the effects of population size and drift, perhaps using approaches such as those reviewed by (Tataru et al., 2017). Furthermore loci may become differentiated between populations not because they are themselves selected but because they are physically close to loci that are selected (Petry, 1983). This phenomenon of linked selection has been the subject of recent research (e.g., (Aeschbacher et al., 2017; Burri, 2017)).

All the results presented here show some dependence on levels of dominance. Although there is interest in the evolution of dominance, little is known of values of levels of dominance in natural populations (Huber et al., 2018; Thurman and Barrett, 2016) except that overdominance is infrequent (Thurman and Barrett, 2016). Here in Figs. 3 and 4 we present evolutionary outcomes for what are, in the absence of overdominance, the extreme values 0.5 and 1. Figs. 3 and 4 suggest that in the absence of overdominance, levels of dominance have only small effects on equilibrium values of F_{ST}.

The method of estimating selection coefficients using equation 5 requires knowledge of migration rates, the allele frequencies in the two habitats and the level of dominance. Migration rates and allele frequencies are routinely measured in genomic studies. While levels of dominance likely have only small effects, it is worth considering whether selection coefficients and the level of dominance at individual loci could be measured directly in genomic field studies. If within each population individuals mate at random, as in the model presented here, then the genotypes of the offspring will be initially at each locus in Hardy-Weinberg frequencies. The offspring are then subject to viability selection which changes the genotype frequencies. Comparison of genotype frequencies in offspring and adults could provide information about selection coefficients and levels of dominance directly.

Suppose, for instance, that at a locus of interest an allele P initially has relative frequency p so that the initial frequencies of offspring genotypes PP, PQ and QQ are as in Table 1, p^2: $2pq$: q^2, then if there are no QQ genotypes in the adults and the ratio of PP to PQ is p^2: $2pq$, it follows that P is dominant and there is complete selection against QQ, i.e., h = 0 and s= - 1 . In general if the ratios of the genotypes in adults are a: b: c then $s = \frac{cp^2}{aq^2} - 1$ and $h = (\frac{b}{2pq} - 1)/s$. In this way selection coefficients and levels of dominance could be measured directly in field studies.

If the strength of selection decreases or migration increases there comes a point at which both alleles are not retained, and one or the other goes to fixation, and F_{ST} declines to zero. This point represents the boundary of what (Sibly and Curnow, submitted) termed the 'polymorphism set' within which stable genetic polymorphisms occur. The polymorphism set is equivalent to the set of non-zero values of F_{ST}, and its boundaries define the limits to local adaptation. The polymorphism set is of particular interest because it can be the starting point for speciation, given a mechanism for phenotype matching and the emergence of a matching cue (Sibly et al., 2019). The boundaries of the polymorphism set were found by (Sibly and Curnow, submitted) by simulation of the evolutionary process, but could alternatively be found by solving the two non-linear simultaneous equations 7.

Here we have derived equations, for populations in migration-selection balance, showing the relationships between equilibrium allele frequencies, selection coefficients, migration rates and dominance. This is illustrated by graphs showing how, at a given locus, F_{ST} increases with the selection coefficients acting on the alleles at the locus, for given migration rates and dominance relationships between the alleles. We hope our methods may help those studying migration-selection balance to quantify the advantages conferred by adaptive genes.

Aeschbacher, S., et al., 2017. Population-genomic inference of the strength and timing of selection against gene flow. Proceedings of the National Academy of Sciences of the United States of America. 114**,** 7061-7066.

Beaumont, M. A., Nichols, R. A., 1996. Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society B-Biological Sciences. 263**,** 1619-1626.

Bulmer, M. G., 1972. Multiple niche polymorphism. American Naturalist. 106**,** 254-257.

Burri, R., 2017. Interpreting differentiation landscapes in the light of long-term linked selection. Evolution Letters. 1**,** 118-131.

Flanagan, S. P., Jones, A. G., 2017. Constraints on the F-ST-Heterozygosity Outlier Approach. Journal of Heredity. 108**,** 561-573.

Graham, A. M., et al., 2018. Migration-Selection Balance Drives Genetic Differentiation in Genes Associated with High-Altitude Function in the Speckled Teal (Anas flavirostris) in the Andes. Genome Biology and Evolution. 10**,** 14-32.

Haldane, J. B. S., 1930. A mathematical theory of natural and artificial selection (Part VI. Isolation). . Proc. Cambr. Philos. Soc. . 26**,** 220-230.

Hoekstra, H. E., et al., 2004. Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. Evolution. 58**,** 1329-1341.

Holsinger, K. E., Weir, B. S., 2009. Genetics in geographically structured populations: defining,

estimating and interpreting FST. Nat Rev Genet. 10**,** 639-650.

Huber, C. D., et al., 2018. Gene expression drives the evolution of dominance. Nature Communications. 9.

Jewett, E. M., et al., 2016. The Effects of Population Size Histories on Estimates of Selection Coefficients from Time-Series Genetic Data. Molecular Biology and Evolution. 33**,** 3002-3027.

Linnen, C. R., Hoekstra, H. E., 2009. Measuring natural selection on genotypes and phenotypes in the wild. Cold Spring Harbor Symposia on Quantitative Biology. 74**,** 155-168.

Lotterhos, K. E., Whitlock, M. C., 2014. Evaluation of demographic history and neutral parameterization on the performance of F-ST outlier tests. Molecular Ecology. 23**,** 2178-2192.

Lotterhos, K. E., Whitlock, M. C., 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology. 24**,** 1031-1046.

Petry, D., 1983. The effect on neutral gene flow of selection at a linked locus. Theoretical Population Biology. 23**,** 300-313.

Rudman, S. M., et al., 2018. What genomic data can reveal about eco-evolutionary dynamics. Nature Ecology & Evolution. 2**,** 9-15.

Savolainen, O., et al., 2013. Ecological genomics of local adaptation. Nature Reviews Genetics. 14**,** 807-820.

Sibly, R. M., Curnow, R. N., submitted. Mapping the selection pressures and dispersal rates that allow local adaptation.

Sibly, R. M., et al., 2019. How phenotypic matching based on neutral mating cues enables speciation in locally adapted populations. Ecology and Evolution. 9**,** 13506-13514.

Tataru, P., et al., 2017. Statistical Inference in the Wright-Fisher Model Using Allele Frequency Data. Systematic Biology. 66**,** E30-E46.

Thurman, T. J., Barrett, R. D. H., 2016. The genetic consequences of selection in natural populations. Molecular Ecology. 25**,** 1429-1448.

Vitalis, R., et al., 2014. Detecting and Measuring Selection from Gene Frequency Data. Genetics. 196**,** 799-+.

Whitlock, M. C., 2011. G '(ST) and D do not replace F-ST. Molecular Ecology. 20**,** 1083-1091.

Wright, S., 1931. Evolution in mendelian populations Genetics. 16**,** 97-159.

Click here to access/download
Supporting File
SI.docx