From inclusive fitness to fixation probability in homogeneous structured populations

Peter D. Taylor*, Troy Day, Geoff Wild[1]

Department of Mathematics and Statistics, Queen's University, Kingston, Ont., Canada K7L 3N6

Abstract

The methods of inclusive fitness provide a powerful analysis of the action of selection on social behaviour. The key component of this analysis is the concept of relatedness R. In infinite populations, a standard method of calculating relatedness coefficients is through coefficients of consanguinity using the notion of genetic identity by descent. In this paper, we show that this approach can also be made to work in finite populations and we assume here that the population has a homogeneous structure, such as an island model. We demonstrate that, under the assumption that genetic effects are small and additive, the resulting formulation of inclusive fitness is equivalent to other significant measures of selection in finite populations, including the change in average allele frequency and fixation probability. The results are illustrated for a model of the evolution of cooperation in a finite island population.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Relatedness; Identity by descent; Genotypic covariance; Cooperation

1. Introduction

Much recent attention has been paid to theoretical aspects of selection in finite structured populations (Rousset and Billiard, 2000; Taylor et al., 2000; Proulx and Day, 2001; Nowak et al., 2004; Wild and Taylor, 2004; Lessard, 2005; Orzack and Hines, 2005). Even more recently, attention has been focused on the evolution of cooperation in a finite population where the population structure is specified with a graph (Lieberman et al., 2005; Ohtsuki et al., 2006; Ohtsuki and Nowak, 2006; Ohtsuki et al., 2007; Taylor et al., 2007). In these studies, a standard measure of the selective advantage of an allele is its fixation probability (i.e. the probability that a single copy of the allele ultimately reaches fixation). However, direct calculation of these fixation probabilities appears to be feasible only when the population has a simple structure such as a cycle; otherwise approximations need to be made (e.g. pair

approximations, Ohtsuki et al., 2006). The inclusive fitness effect of an allele (Hamilton, 1964) can be exactly calculated in more complex structures, however, and if we are prepared to assume that selection is weak (an approximation of another kind) then this provides an accurate measure of relative fixation probability. This was shown by Rousset and Billiard (2000) and our purpose here is to provide an alternative approach to this important result with slightly different assumptions (e.g. describing population dynamics with the Moran model). As with any inclusive fitness calculation, the cornerstone of our approach is a careful formulation of relatedness (Michod and Hamilton, 1980; Seger, 1981; Grafen, 1985; Rousset and Billiard, 2000; Rousset, 2002; Grafen, 2006). Our main result is that, under appropriate assumptions, relatedness in finite structured populations can be calculated using the concept of identity by descent. Moreover, the inclusive fitness effect formulated using this relatedness then provides an equivalent description of evolution in finite populations to fixation probability.

For simplicity our results are formulated within a homogenous population, that is, one in which each individual "sees" the same structure, where structure refers

---

*Corresponding author. Tel.: +1 613 533 2428; fax: +1 613 533 2964.

E-mail addresses: peter.taylor@queensu.ca (P.D. Taylor), tday@mast.queensu.ca (T. Day), gwild@uwo.ca (G. Wild).

[1]Current address: Department of Applied Mathematics, The University of Western Ontario, London, Ont., Canada N6A 5B7.

Nomenclature

B, $\bar{B}$ individual birth rate, population-wide average rate

D, $\bar{D}$ individual death rate, population-wide average rate

BD, DB   birth–death, death–birth protocols (the Moran model—Section 4)

δ phenotypic penetrance of allele A

d_{ij} probability an offspring of i displaces individual j.

e_{ij} probability i will interact with j.

ε probability of a birth in the population in a small time interval

E( ) & cov(,)   expectation and covariance over the population in a fixed state

E[·] & cov[,]   expectation and covariance for a finite population with respect to the equilibrium distribution defined over all population states

G probability that two alleles are identical by descent (IBD), i.e. the coefficient of consanguinity between two given individuals

G_i coefficient of consanguinity between actor and ith interactant

G^* coefficient of consanguinity between two individuals chosen at random from the population (with replacement)

G_0 coefficient of consanguinity between an individual and itself (for haploid individual $G_0 = 1$)

s in a deme-structured population, probability a breeder is native to the deme. Is also the

probability an offspring displaces an individual in its native deme

m deme size

n number of demes in the population

N population size

$p = u_1/u$ probability with which a mutation at birth results in an A-allele under the contrived mutation process (implicitly, the equilibrium frequency of A in the absence of selection)

Q state transition matrix

r the probability an individual's next interaction is with its most recent offspring

R coefficient of relatedness

ρ_A fixation probability of a single copy of allele A.

T_i the proportion of time at equilibrium a finite population spends in state i

u_1, u_2 the rate (or effective rate, under the contrived process) of mutation from A to B, and from B to A, respectively

$u = u_1 + u_2$ the symmetric rate of mutation under the contrived mutation process

W individual fitness

W_X first partial derivatives of W evaluated at $\delta = 0$

W_I the inclusive fitness effect

x, y, $x\sim$, $y\sim$ within-individual allele frequencies (genotypic values)

$\bar{x}$, $\bar{y}$ population-wide average value of x and y, respectively (in a finite population this depends on the population state)

X, X_i phenotypic values

to the flow of individuals (the pattern of dispersal), and the network of fitness-determining interactions. An example is a deme-structured population, finite or infinite, with equal sized demes and a symmetric structure within and between demes (Fig. 1). Internal asymmetries (e.g. demes of different size) can be handled with our methods, but require the use of relative reproductive values (e.g. Taylor and Frank, 1996) and we leave that for future work.

Section 2 presents some crucial technical results for calculating covariances that will be necessary for specifying relatedness correctly. Section 3 then formulates the inclusive fitness effect of an allele, and specifies relatedness using the results of Section 2. For the sake of comparison and completeness, we present results for finite and for infinite populations in both of these sections. In Section 4, we then show that, under the assumption of weak selection, the inclusive fitness effect derived in Section 3 predicts relative fixation probability in a finite structured population. Section 5 then provides an example of the selective advantage of cooperation in an island model.

2. Some preliminary technical results

As with many previous inclusive fitness analyses, our treatment will be based on Price's (1970) covariance formulation of selection. This requires that one calculate the genotypic covariance between various individuals within the population. To do this directly we need to get hold of the underlying allelic distribution and this can be difficult. A standard alternative approach uses the notion of identity by descent.

Suppose that at the locus of interest there are two alleles A and B (called the two *states* of the gene) with per generation mutation rates u_1 from B to A and u_2 from A to B. We let an individual's genotypic value x be the frequency of A in its genotype, and we seek a formula for cov(x, y), the covariance between the genotypic value of two interacting individuals. This is often computationally feasible only if the alleles are neutral (phenotypically identical) and we assume that now.

We say that two genes at the same locus are *identical by descent* (IBD) if they both descend from the same mutational event with no intervening mutation, and we
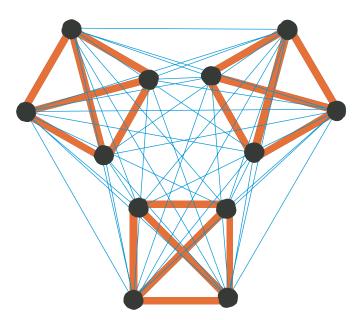
Fig. 1. An example of Wright's (1943) island model represented as a graph with nodes representing individuals and edges representing interactions between individuals. These interactions can be of two types, describing fitness interactions and dispersal patterns of offspring. In this case there are $N = 12$ individuals divided up into $m = 3$ demes of size $n = 4$.

define the *coefficient of consanguinity G* between two individuals to be the probability that two genes chosen at random, one from each individual, are IBD. We begin with a common but dubious argument relating cov(x, y) to G.

Take two interactants with genotypic values x and y, and choose a gene at random from each. Either the genes are IBD (probability G) or they are not (probability $1-G$). If they are, then $x = y$ and the covariance is simply the variance of x (assuming x and y have the same distribution). If they are not, they are independent and have covariance 0. Thus

$$\text{cov}(x, y) = G \, \text{var}(x) + (1 - G)0 = G \, \text{var}(x). \quad (2.1)$$

There are two places where this argument makes dubious assumptions. The first is the assumption that if the genes are IBD then the covariance is the variance of x. If we are given the condition that two genes are IBD, are their conditional allele frequencies the same as the population-wide frequencies? In fact, often they are not. The second is the assumption that genes that are not IBD are independent. This is also false in general, even in an infinite population. For example, take an infinite island-structured population with vanishingly little dispersal between islands, so that almost all local genetic variation comes from mutation which is itself rare. Then almost all islands will be found in a state of fixation, release from which will generally come with a mutation event. In that case, if two alleles from the same island are not IBD, they are almost certain to be different in state and therefore not independent.

Though the argument yielding Eq. (2.1) is flawed, it is a formula we would really like to be able to use, as

coefficients of consanguinity (in a homogeneous structured population) can typically be calculated with recursive methods. Fortunately, the argument for Eq. (2.1) can be fixed with a careful organizational change.

Whenever we calculate a covariance there is an assumed underlying probability distribution, and the first job is to specify that. We will define what we call the equilibrium distribution, and the precise description of this distribution will be different in a finite and infinite population.

In the infinite population the distribution of alleles in the population generally settles down to a time-invariant distribution determined by a balance between mutation, selection, and offspring dispersal. In that case, we take this to be the equilibrium distribution.

The finite population analogue of this deserves a bit more care. Let us define the *state* of the population to be the distribution of alleles among the individuals. Under the forces of selection, drift and mutation, this distribution of alleles does not settle down to an asymptotic equilibrium as in the infinite population case; rather the population continually fluctuates among different states. To get an "overall" measure, we presumably want some sort of average over all possible states, and the key question then is how to appropriately weight the different states in this calculation. We adopt here what is perhaps the most obvious choice and weight each population state by the relative amount of time the population spends in that state if it is allowed to evolve for a very long-time. This particular weighting was introduced by Rousset and Billiard (2000) and was subsequently employed by Taylor et al. (2000). We take this as the equilibrium distribution for a finite population—the set of population states endowed with the probability distribution determined by letting a single population run for an infinite time.

Now when we come to calculate expectations or covariances among random variables defined in a finite population, it will be important to clearly distinguish whether we are working within a single population state or across all states endowed with this equilibrium distribution. To signal this, we will use round brackets for the former and square brackets for the latter. For example, in an island model (Fig. 1), if x is the genotypic value of a random individual in the population and y is the genotypic value of a random deme-mate, then E(x) denotes the expectation taken over all individuals in a particular population state and cov(x, y) denotes covariance calculated over all pairs of deme-mates, again in a particular state. Both of these quantities depend on the state. It will be convenient to use the abbreviation $\bar{x} = \text{E}(x)$ for the average genotypic value (also the average allele frequency) in the population state. On the other hand, $\text{E}[x] = \text{E}[\bar{x}]$ denotes the expectation of x over all states and thus it is the long-term average allele frequency. Similarly cov[x, y] denotes the long-term genotypic covariance between deme-mates.

To emphasize this key distinction, a standard result (Ross, 1998) decomposes this long-term covariance as the

sum of a "within-state" and a "between-state" component:

$$\text{cov}[x, y] = \text{E}[\text{cov}(x, y)] + \text{cov}[\text{E}(x), \text{E}(y)]. \tag{2.2}$$

This decomposition plays a critical role in the proof of Proposition 2.1 below. The reason for this is that selection happens within the population at a particular moment, and for this reason, the genotypic covariances we use to measure selection (e.g. in Price's (1970) formula) must be taken within a state (Seger, 1981). To get an overall average covariance, we then take the expectation of these over all states, and that is the first term on the right in Eq. (2.2). On the other hand, to get a valid argument for Eq. (2.1), the covariance we need to work with turns out to be the long-term measure found on the left in Eq. (2.2). The difficulty in the argument (and much of the confusion!) comes from the fact that these two covariance measures are not the same.

The other "fix" we need is a technical modification of the mutation process. Suppose that alleles A and B each mutate at rate $u = u_1 + u_2$ and that they each mutate to A with probability $p = u_1/u$ and to B with probability $1 - p = u_2/u$. We will call this new process the *contrived* mutation process. Under this process the effective rate at which B mutates to A is $pu = u_1$ and the effective rate at which A mutates to B is $(1 - p)u = u_2$, *so that as far as the state of a gene is concerned, there is no distinction between the population with the original mutation process and one with this contrived process.* In particular, this change in the mechanism of mutation has no effect on the covariance of genotypic values. The only difference is that, under this higher contrived mutation rate, two genes are less likely to be IBD than before because, for example, when A mutates to A, we will no longer consider the new gene to be IBD to the old. *Henceforth our notion of IBD will assume this contrived mutation rate.* Our adoption of this new mutation regime gives us the following key result which we prove in the Appendix.

Proposition 2.1. *Take the equilibrium distribution for both the infinite and the finite population with two neutral alleles A and B and the contrived mutation rate described above. Suppose that "two" individuals (could be the same individual twice), with genotypic values x and y, are chosen with respect to the population structure, and take a random gene from each. If the two genes are IBD then they are both A or B with probability p or 1–p, respectively; if they are not IBD, then they come from different mutational events and they are each (independently) A or B with probability p or 1–p, respectively. It follows from this that*

Infinite population : $\text{cov}(x, y) = G\bar{x}(1 - \bar{x}).$ \hfill (2.3)

Finite population : $\text{E}[\text{cov}(x, y)] = (G - G^*)\text{E}[\bar{x}](1 - \text{E}[\bar{x}]).$ \hfill (2.4)

Here G is the coefficient of consanguinity between the individuals and G^* is the coefficient of consanguinity of two randomly chosen individuals in the population,

allowing for the possibility that the same individual is chosen both times (sampling with replacement).

In Eq. (2.3), the covariance is over the population and in Eq. (2.4) the covariance is conditional on the state and E[] denotes the average over all states. In both cases, $\bar{x}$ is the average value of x in the population; in Eq. (2.4) this average is conditional on the state. It is important to emphasize that, in this Proposition, when two individuals are chosen, they are chosen at the same moment (therefore from the same state), but they could be taken at any time (therefore from any state) according to the weighting provided by the equilibrium distribution.

3. Inclusive fitness calculation

The previous section provides some crucial technical results that will now be used to properly specify relatedness and formulate the inclusive fitness effect of a mutant allele A. Rather than assume neutrality, as in the previous section, we now let selection act at a weak level. In particular, we suppose that two "neighbours" interact, and the behaviour of each is given by its phenotypic value X which in turn is correlated with its genotypic value x. We let A have a small effect δ on the behaviour of an individual in the following way: we let the allele B have genic value 0 and the allele A have genic value 1, and suppose that an individual with genotypic value x has infinitesimal phenotypic change $dX = \delta x$. Under diploidy, this effect would be called additive.

We assume that population dynamics follow the well-known Moran process (Moran, 1958, 1962) with fecundity selection. Thus the fecundity of an individual determines its probability of producing an offspring. We let individual fecundity depend on the genotype of the individual and of certain neighbours (determined by the population structure) whose behaviour affects the fitness of the individual. In a differential time interval dt, the probability that an individual gives birth will have the form Bdt, and the probability that an individual is displaced by an offspring will have the form Ddt, where, in general, B and D depend on the fecundity of the individual and of others in the population. Mathematically we regard them as differentiable functions of the phenotypes X_i of all interactants (including the individual itself): $B = B(X_1, X_2, \ldots, X_n)$ and $D = D(X_1, X_2, \ldots, X_n)$ but we will generally suppress these variables. Since every birth results in a death, $\bar{B} = \bar{D}$.

We now calculate the effect of the deviant behaviour on the population-wide allele frequency in a single population (finite or infinite). In differential time dt the change in the average allele frequency will be

$$d\bar{x} = \text{E}[Bdt((1 - u)x + up) - Ddtx], \tag{3.1}$$

$$d\bar{x} = \text{E}(xW)dt - u[\text{E}(xB) - p\bar{B}]dt, \tag{3.2}$$

where we take fitness to be $W = B - D$. In Eq. (3.1) the birth term for an x-individual is split into two pieces; when there

is no mutation (probability 1–u) the offspring allele is A with probability x; otherwise (probability u) it is A with probability p. Rearranging Eq. (3.2):

$$\mathrm{d}\bar{x}/\mathrm{d}t = \mathrm{cov}(x, W) - u[\mathrm{cov}(x, B) + (\bar{x} - p)\bar{B}] \qquad (3.3)$$

using the fact that $\bar{W} = 0$. Not unexpectedly, Eq. (3.3) has the form of the covariance formula of Price (1970). Price pointed out the need for a "second" term in cases such as non-random meiosis and fertilization in which the allele frequency among gametes differs from the parental genotype. Mutation has exactly this effect and accounts for the final term in Eq. (3.3).

Now we write fitness, to first order in δ, in a Taylor expansion in terms of the genotypic values x_i:

$$W = 1 + \sum_i W_{X_i}\mathrm{d}X_i = 1 + \delta \sum_i W_{X_i}x_i + o(\delta). \qquad (3.4)$$

Here the subscripts on W denote partial derivatives, and these are evaluated at $\delta = 0$, and we take mean fitness in a neutral population to be 1. The "weak selection" assumption that δ is small is standard for inclusive fitness arguments (Hamilton, 1964; Charnov, 1977) and the results we obtain are valid to first order in δ. Now putting Eq. (3.4) into Eq. (3.3), we get

$$\mathrm{d}\bar{x}/\mathrm{d}t = \delta \sum_i W_{X_i}\mathrm{cov}(x, x_i)$$
$$\qquad - u[\mathrm{cov}(x, B) + (\bar{x} - p)\bar{B}] + o(\delta). \qquad (3.5)$$

Eq. (3.5) describes allele frequency change in a single population. In the infinite population case, that is all we have, but in a finite population we will want to take an expectation over the equilibrium distribution of all population states. Henceforth we treat these two cases separately.

Infinite population: In a neutral population, genotype has no effect on fitness, so that $\delta = 0 = \mathrm{cov}(x, B)$, and if we set $\mathrm{d}\bar{x}/\mathrm{d}t = 0$ in Eq. (3.5), we get $\bar{x} = P$; at equilibrium, the neutral allele frequency is determined by the relative mutation rates. When δ is increased above zero and selection is allowed to act, we generally take the mutation rate u to be much smaller than δ, so that its effects are negligible beside those of selection. We then write Eq. (3.5) as

$$\mathrm{d}\bar{x}/\mathrm{d}t = \delta\,\mathrm{var}(x)W_{\mathrm{I}} + o(\delta) + O(u), \qquad (3.6)$$

where

$$W_{\mathrm{I}} = \sum_i W_{X_i}R_i \qquad (3.7)$$

is the inclusive fitness effect of the interaction, and R_i is the relatedness of the focal individual to the ith interactant:

Infinite pop. $R_i = \dfrac{\mathrm{cov}(x, x_i)}{\mathrm{var}(x)} = \dfrac{G_i}{G_0}. \qquad (3.8)$

In Eq. (3.8) we use the neutral ($\delta = 0$) distribution to calculate the genotypic covariances. The error introduced will be of order δ, and when this is inserted into the δ-term in Eq. (3.6), it can be absorbed into the $o(\delta)$ term at the

end. We have also used Eq. (2.3) to write R_i in terms of coefficients of consanguinity, writing G_0 as the coefficient of the focal individual with itself. Eq. (3.6) gives us the following result.

Proposition 3.1. *In an infinite population at the neutral equilibrium, with negligible mutation rates, the inclusive fitness W_{I} predicts the direction of change of mutant allele frequency when the action δ of this allele is increased above zero.*

Finite population: Eq. (3.5) gives us the rate of change of the population-wide allele frequency, and in a finite population this will depend on the particular population state. Following Eq. (2.4), we take the expectation of Eq. (3.5) over all population states:

$$\mathrm{E}[\mathrm{d}\bar{x}/\mathrm{d}t] = \delta \sum_i W_{X_i}\mathrm{E}[\mathrm{cov}(x, x_i)]$$
$$\qquad - u\mathrm{E}[(\bar{x} - p)\bar{B}] + o(\delta) + o(u), \qquad (3.9)$$

$$\mathrm{E}[\mathrm{d}\bar{x}/\mathrm{d}t] = \delta \sum_i W_{X_i}\mathrm{E}[\mathrm{cov}(x, x_i)] - u(\mathrm{E}[\bar{x}] - p)\mathrm{E}[\bar{B}]$$
$$\qquad + o(\delta) + o(u). \qquad (3.10)$$

In Eq. (3.9) we have used the fact that $\mathrm{cov}(x, B)$ goes to zero as w approaches zero, because for small u the equilibrium distribution is almost entirely concentrated at the two fixation states at which the birth rate B is the same for all individuals. Thus this term in Eq. (3.5) is absorbed into the $o(u)$ term in Eq. (3.9). In Eq. (3.10) we have used the fact that $\mathrm{cov}[(\bar{x} - p), \bar{B}] = 0$ since the average birth rate $\bar{B}$ is the same in all population states.

As in Eq. (3.6), we write Eq. (3.10) as

$$\mathrm{E}[\mathrm{d}\bar{x}/\mathrm{d}t] = \delta\mathrm{E}[\mathrm{var}(x)]W_{\mathrm{I}} - u(\mathrm{E}[\bar{x}] - p)\mathrm{E}[\bar{B}]$$
$$\qquad + o(\delta) + o(u), \qquad (3.11)$$

where as in Eq. (3.7), the inclusive fitness effect is

$$W_{\mathrm{I}} = \sum_i W_{X_i}R_i,$$

where the relatedness between interactants has the form:

Finite pop. $R_i = \dfrac{\mathrm{E}[\mathrm{cov}(x, x_i)]}{\mathrm{E}[\mathrm{var}(x)]} = \dfrac{G_i - G^*}{G_0 - G^*}. \qquad (3.12)$

As in Eq. (3.8) the covariances are calculated using the neutral ($\delta = 0$) distribution. Note that the infinite population relatedness (Eq. (3.8)) can be regarded as a special case of Eq. (3.12) because, in an infinite population, $G^* = 0$. As a check on Eq. (3.12) note that if recipient i is a random member of a finite population (who could be the actor with probability $1/N$) then $G_i = G^*$ and hence $R_i = 0$. Average relatedness in a finite population is zero. This is expected— a gift of fitness which is spread out evenly over the entire population cannot, on average, be of net benefit to any individual.

For a finite population, the equilibrium distribution is a long-term steady state, and $\mathrm{E}[\mathrm{d}\bar{x}/\mathrm{d}t]$, the average rate of increase of the allele frequency, must be zero. Setting this

to be zero in Eq. (3.11), and taking the limit as u approaches zero, we get

$$\delta W_I = K(E[\bar{x}] - p) + o(\delta), \tag{3.13}$$

where $K = E[\bar{B}][\lim_{u \to 0} E[\text{var}(x)]/u]^{-1}$. The simplest way to show that this limit exists and is finite is to jump ahead and refer to the eigenvector analysis in Section 4. Using row 2 and row 3 of the matrix in Eq. (4.1), we argue that at equilibrium, $T_{0+} = O(u_1)$ and $T_{N-} = O(u_2)$. It follows that the probability of being unfixed is $O(u)$. Since $\text{var}(x) = 0$ in either of the fixation states, we deduce that $E[\text{var}(x)] = O(u)$.

As a check on Eq. (3.13) note that with neutral alleles ($\delta = 0$) the equilibrium allele frequency is $E[\bar{x}] = p$, as expected. Eq. (3.13) tells us that the inclusive fitness effect W_I has the same sign as $E[\bar{x}] - p$. This is an IBD analogue of a result of Rousset and Billiard (2000, Eq. 15) and we state it as a proposition.

Proposition 3.2. *In a finite population, for sufficiently small additive selective effects $\delta > 0$, the inclusive fitness effect W_I will predict the side of the neutral equilibrium on which the new equilibrium (determined by the selection–mutation balance) will occur.*

Numerical simulations that we have run in small deme-structured populations illustrate this (Taylor et al., 2000).

4. The connection between inclusive fitness and fixation probability

Here we tie our analysis of inclusive fitness in a finite structured population to fixation probability. For simplicity, we work with an asexual haploid population of fixed size N, with two alleles A and B. Population dynamics are stochastic and follow the Moran process as described in Section 3.

To simplify the analysis, we reduce the state space of the population to four population states. States 0 and N will be the fixation states of "all B" and "all A", respectively. The remaining two states we call $0+$ and $N-$, defined as follows. When the population leaves state 0, we assign it to state $0+$ where it remains until it hits either state 0 or state N. Similarly when the population leaves state N, we assign it to state $N-$ where it remains until it hits either state 0 or state N. Thus, the states $0+$ and $N-$ are made up of many different sub-states of the state space for the original model of interest. With the ordering of these new states $[0, 0+, N-, N]$, the state transition matrix corresponding to a small time interval has the form:

$$Q = \begin{bmatrix} 1 - u_1\varepsilon & h(1 - \rho_A) & k\rho_B & 0 \\ u_1\varepsilon & 1 - h & 0 & 0 \\ 0 & 0 & 1 - k & u_2\varepsilon \\ 0 & h\rho_A & k(1 - \rho_B) & 1 - u_2\varepsilon \end{bmatrix}, \tag{4.1}$$

where entry q_{ij} gives the probability of moving from state j to i.

The transition matrix (4.1) can be derived as follows. First, if we denote the probability that there will be a birth in the given small time interval by ε (which will be proportional to dt), then columns 1 and 4 follow from the definition of u_1 and u_2. Next consider column 2. Define ρ_A as the probability that beginning with a single A-bearing individual the population hits state N before it hits state 0. Similarly define ρ_B as the probability that beginning with a single B-bearing individual the population hits state 0 before it hits N. Note that, if $u = 0$, these are the standard fixation probabilities. Next we define h as the probability that the population leaves state $0+$ in the next move. This quantity will depend on the distribution among the many sub-states making up state $0+$, and we take this distribution to be the population equilibrium distribution described in Section 2. Entry $q_{1,2}$ should then be h multiplied by the probability that the system hits state 0 (as opposed to state N) at the point of exit from state $0+$. Similarly, entry $q_{4,2}$ should be h multiplied by the probability of hitting state N, at the point of exit from state $0+$. Now a standard result says that the ratio of the transition probabilities $q_{4,2}$ to $q_{1,2}$ is the same as the ratio of the ultimate probabilities that $0+$ will move to N or to 0. We know that this ratio of ultimate probabilities is $\rho_A/(1-\rho_A)$, and therefore the ratio of $q_{4,2}$ to $q_{1,2}$ must equal $\rho_A/(1-\rho_A)$. This gives column 2. Column 3 can be derived analogously, defining k as the probability that the population leaves state $N-$ in the next move.

With transition matrix (4.1) we can now proceed by defining T_i to be the long-term proportion of time the population spends in state i. The equilibrium state distribution vector $[T_i]$ is the right eigenvector of Q for the eigenvalue 1. It follows easily from the eigenvector equations that

$$\frac{T_N/u_1}{T_0/u_2} = \frac{\rho_A}{\rho_B}. \tag{4.2}$$

Take note that both the T's and the ρ's in Eq. (4.2) depend on the mutation rate, and we are primarily interested in the values of ρ_A and ρ_B in the limit as the mutation rate, u, approaches zero. In this limit, the population will be entirely concentrated in the two fixation states, with $T_N = E[\bar{x}]$, and $T_0 = 1 - E[\bar{x}]$. Since $u_1/u_2 = p/(1-p)$, we can conclude that, in the $u = 0$ limit, $E[\bar{x}] > p$ precisely when $\rho_A > \rho_B$.

Putting this together with Proposition 3.2, we get

Proposition 4.1. *If selection is weak and gene action is additive (δ is small enough), then in the limit as mutation rate u approaches zero, the three conditions below are equivalent:*

$$\rho_A > \rho_B, \tag{4.3}$$

$$E[\bar{x}] > p, \tag{4.4}$$

$$W_I > 0. \tag{4.5}$$

This provides an alternative route to an important result of Rousset and Billiard (2000).

5. Application to cooperation in homogeneous graphs

A striking application of the general theory is found in a recent studies of cooperative behaviour in an asexual haploid population in which individuals are represented as inhabiting the nodes of a finite graph satisfying a general homogeneity condition (Taylor et al., 2007, Grafen and Archetti, in preparation). For any two individuals, i and j, let d_{ij} be the probability that the next offspring of i will replace j (the population size remains constant) and let e_{ij} be the probability that the next fitness interaction of i will be with j, and assume that these are symmetric, $d_{ij} = d_{ji}$ and $e_{ij} = e_{ji}$. We suppose that $d_{ii} = e_{ii} = 0$, and note that $\sum_j d_{ij} = \sum_j e_{ij} = 1$. An *isomorphism* of the graph is a permutation of the nodes that preserves the edge weights d_{ij} and e_{ij}. The graph is called *transitive* if given any pair of nodes i and j, there is an isomorphism mapping i to j. Roughly speaking, a graph is transitive if it "looks the same" from any node. Two common examples of such graphs are found in finite stepping-stone populations (cycles) and finite island models. Taylor et al. (2007) use an inclusive fitness analysis to obtain simple general conditions for the fixation of an altruistic allele in such a population under a Moran model (Section 3). They assume a slightly stronger condition, bi-transitivity, but Grafen and Archetti (in preparation) extends their results to transitive graphs (and also remove the assumption that $d_{ii} = e_{ii} = 0$).

We let the fecundity f_i of individual i be determined through fitness interactions with neighbours (determined by the e_{ij}) and we consider two versions of fecundity selection. In the BD process, births are allocated to the population at a fixed, fitness-independent rate and are given to individual i with relative probability f_i, replacing a neighbour j of i with probability d_{ij}. In the DB process, individuals die at a fixed, fitness-independent, rate and a death at node j is replaced by an offspring from node i with relative probability $f_i d_{ij}$. Take note that the f_i measure fecundity, but this is only one component of fitness, the other being mortality (from replacement).

Let an altruist give fecundity benefit b to an interactant at cost c. Then, in a population of size N, the inclusive fitness condition (Eq. (4.5)) that the altruistic allele be favoured is

$$\text{BD protocol}: c < -\frac{1}{N-1}b, \tag{5.1}$$

$$\text{DB protocol}: c < \frac{Nr-2}{N-2}b, \tag{5.2}$$

where $r = \sum_j d_{ij} e_{ij}$ and is (by the transitive condition) independent of i. The parameter r can be interpreted as the average d—weight of the edge from the focal individual to its next interactant. It is surprising that condition (Eq. (5.1)) is independent of both the offspring dispersal probabilities d_{ij} and the interaction probabilities e_{ij}. It is therefore independent of the population structure (provided it is transitive) and is the same as for a random mixing population ($d_{ij} = e_{ij} = 1/(N-1)$). This provides an extension of an infinite-population result of Wilson et al. (1992) and Taylor (1992a,b). Condition (5.2) has the familiar form of a condition for cooperation; it says that the cost must be small and the benefit large. Condition (5.1) however, tells us that under the BD protocol, conventional forms of cooperation can never evolve. The condition can hold with a positive benefit if c is negative (giving a benefit to the actor), a form of mutualism, or with positive cost if b is negative (doing harm to the recipient), a form of spite. It is remarkable that this condition is so general, applying in a large number of homogeneous structured populations in which direct calculation of fixation probabilities would be infeasible.

As a particular example, consider an island model (Fig. 1) with n demes each with m nodes, so that $N = nm$. Let an offspring stay in its native deme with probability s and migrate to another deme with probability $1-s$. This gives us edge weights $d_{ij} = s/(m-1)$ within a deme. If interactions are at random within each deme, so that $e_{ij} = 1/(m-1)$ within a deme, then $r = s/(m-1)$. If we set $s = 3/4$ and $m = 4$, then $r = 1/4$ and condition (5.2) can be rewritten as

$$b > \frac{N-2}{n-2}c. \tag{5.3}$$

If we let population size increase, keeping deme size and migration probability fixed, we get

$$n = 3 \text{ demes}(N = 12) \quad b > 10c,$$
$$n = 8 \text{ demes}(N = 32) \quad b > 5c,$$
$$n = \infty \text{ demes}(N = \infty) \quad b > 4c.$$

As the population grows, cooperation is increasingly favoured.

6. Discussion

The purpose of this paper is to provide, for a finite population, a rigorous foundation for Proposition 4.1 which, under an assumption of weak selection and additive gene effects, relates standard notions of selective advantage, fixation probability and average allele frequency to Hamilton's (1964) notion of inclusive fitness. This result has great practical significance, as in many cases inclusive fitness is much easier to calculate and to work with than are fixation probability and average allele frequency. It is interesting to note that a direct link between fixation probability (Eq. (4.3)) and inclusive fitness effect (Eq. (4.5)) is hard to obtain; the connection works through average allele frequency change (Eq. (4.4)).

In our development of the ideas, we have discussed both finite and infinite populations, partly for completeness

sake, and partly to allow a comparison of concepts such as relatedness. In both cases weak selection and additivity of gene action and phenotypic effects on fitness are key assumptions. In an infinite population, the inclusive fitness W_I predicts the direction of change of mutant allele frequency when the action of this allele is altered from the neutral behaviour (Proposition 3.1). In a finite population, we need to work with the balance between mutation and selection, and the inclusive fitness W_I predicts the side of the neutral equilibrium on which the new selective equilibrium will occur (Proposition 3.2).

Identity by descent: Relatedness is defined in terms of genotypic covariance and the calculation of these covariances requires knowledge of the allelic distribution. In a structured population this can be difficult to get hold of. An elegant way around this problem is to condition on the IBD status of the alleles (Eq. (2.1)). For this approach to work, we want IBD alleles to have the population-wide allele frequency and non-IBD alleles to be independent. Unfortunately, neither of these conditions generally holds. This has led some authors to suggest alternative IBD notions that *guarantee* that equations like Eq. (2.1) hold (Rousset (2002) and references therein). The coefficient G presented in this paper can be placed among these "alternative" IBD measures. Establishing a proper theoretical basis for this requires a bit of careful technical work and this is the mandate of Section 2. A key step is the introduction of the contrived mutation rate. This reformulation has no effect on allelic states, but alters only the question of when two genes are to be considered IBD. In this sense, it is a purely technical device to allow us to prove rigorous theorems that connect relatedness to identity by descent.

In Section 3, we convert Price's equation for allele frequency change into an inclusive fitness formulation using relatedness coefficients, which turn out to be quotients of covariances (the middle terms in Eqs. (3.8) and (3.12)). In order to be able to express these in terms of coefficients of consanguinity, we need the technical work of Section 2.

The role of mutation: In many ways in this analysis, mutation itself plays a purely technical role. In an infinite population you can often do without mutation altogether. In a finite population, it serves two essential purposes. First, with no mutation, all coefficients of consanguinity are 1 and relatedness (Eq. (3.12)) is of the form 0/0. The purpose of mutation is to resolve the 0/0 limit, but in the end, it cancels out (to first order) and our expression for R is independent of u. Secondly, with no mutation, the population will eventually become fixed for one allele or the other and selection cannot act. Thus mutation is essential to provide the equilibrium distribution. However, a very small mutation rate is sufficient to do the job, and indeed our main result, Proposition 4.1, makes sense in the limit as u approaches zero. [Of the main entities in Proposition 4.1, the inclusive fitness and the fixation probabilities are defined without

mutation, and the equilibrium allele frequency is defined in the zero mutation limit.]

Weak selection and synergistic effects: Note that the assumption that fitness can be expanded in a Taylor series in terms of phenotypic deviations (Eq. (3.4)) implies that synergistic effects, arising when both interactants are deviant (Queller, 1985), are second order in δ and can be ignored. In that sense, additivity of gene action is a consequence of our assumption of weak selection. To clarify this further, there are two forms of weak selection in the literature the one we use here assumes that alleles have a small effect on behaviour, and therefore on fitness; the alternative notion, e.g. Nowak et al. (2004), allows alleles to have a large effect on behaviour, but assumes that these behavioural changes have a small effect on fitness. Under our assumption, when two deviant individuals interact, the fitness effect on each is (to first order in δ) the sum of the effects of each deviation; under the alternative assumption, this is not the case and we must account for possible synergistic fitness effects in the interaction (Queller, 1985). As an example, suppose that interactions can be modelled as a 2-person game with matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, where a_{ij} is the incremental payoff to strategy i against strategy j, and let us suppose that the resident is 1 and the deviant is 2. Weak selection in the alternative sense requires only that all entries of the matrix be small (Nowak et al., 2004), but weak selection in our sense requires that the fitness effect of the interaction of two deviants be the sum of the individual effects, this is, that $a_{22} - a_{11} = (a_{12} - a_{11}) + (a_{21} - a_{11})$, or more precisely, that $a_{11} + a_{22} - a_{12} - a_{21}$ is $o(\delta)$. For example, this is the case for the cooperative behaviour example of Section 6, which can be modelled with the matrix $\begin{bmatrix} b - c & -c \\ b & 0 \end{bmatrix}$. It can be shown (e.g. Wild and Traulsen, 2007; Taylor et al. in prep.) that the fixation probability condition (4.3) that $\rho_A > \rho_B$ is equivalent to the condition $\rho_A > 1/N > \rho_B$ under our assumption of weak selection which implies the matrix game is additive, but not otherwise. For example, even with small a_{ij}, one can have ρ_A and ρ_B both greater than $1/N$ or both less than $1/N$ (Nowak et al., 2004; Ohtsuki et al., 2007).

Appendix A

A.1. Proof of Proposition 2.1

The argument for this proposition hinges on the observation that every local configuration occurs many times, either in different corners of the infinite population or at different instants in the (infinite) life of the finite population, and a mutational event can be thought of as happening in many of these corners (infinite pop) or at many different instants (finite pop) and a proportion p of these will result in an A allele and a proportion $1–p$ of these will result in a B allele. For example, suppose that two A-genes are IBD. Then the mutational event which created their common ancestor occurred either in a corner (infinite pop), or at an instant (finite pop) for which in a large number of identical corners or instants the same mutation event produced the two alleles A and B with probabilities p and $1–p$. A similar argument shows that if two genes are not IBD, then each of them will be paired, in that situation, with A or B with probability p or $1–p$, respectively. For this argument to work it is crucial that the alleles be neutral and are therefore under identical forces (of mutation and drift) so that the potential trajectory of a gene is independent of its state.

Now take an infinite population. First we choose random genes $x\sim$ and $y\sim$ from the two individuals. Now either $x\sim$ and $y\sim$ are IBD (probability G) or they are not (probability $1–G$). It follows from the argument above that, in the first case, they are both A with probability p and both B with probability $1–p$, for a covariance of $p(1–p)$. In the second case they are independent, giving a covariance of 0. Thus

$$\text{cov}(x, y) = \text{cov}(x\sim, y\sim)$$
$$= Gp(1 - p) + (1 - G)0 = G\bar{x}(1 - \bar{x}), \qquad (A.1)$$

where we have used the fact that at equilibrium, average allele frequency is $\bar{x} = p$. Note that for the second equality we have used the covariance decomposition principle with two classes, the class in which $x\sim$ and $y\sim$ are IBD and the class in which they are not. Over both classes, $x\sim$ and $y\sim$ have mean $\bar{x}$ so the covariance of the class averages is zero, and Eq. (A.1) displays the average within-class covariance only. Finally, to obtain the first equality, under haploidy this is trivial, and under diploidy this again uses the covariance decomposition theorem where the classes are pairs of individuals and x and y are the class averages. The within-class covariance is always zero as $x\sim$ and $y\sim$ are chosen at random in each individual. This gives us Eq. (2.3).

In the finite population, the argument is exactly the same provided we use the equilibrium distribution on the space of all states of the population. We get

$$\text{cov}[x, y] = G\text{E}[\bar{x}](1 - \text{E}[\bar{x}]). \qquad (A.2)$$

The reason for this is that what we needed to make the infinite population argument work was that every possible allelic configuration (from all possible mutational and parental sampling trajectories) be represented with its appropriate probability and to get the same situation in a finite population we need the full set of possible population states, each with its appropriate frequency.

Now recall Eq. (2.2):

$$\text{cov}[x, y] = \text{E}[\text{cov}(x, y)] + \text{cov}[\bar{x}, \bar{y}].$$

If we eliminate $\text{cov}[x, y]$ from (A.2) and Eq. (2.2), we get

$$\text{E}[\text{cov}(x, y)] = G\text{E}[x](1 - \text{E}[x]) - \text{cov}[\bar{x}, \bar{y}]. \qquad (A.3)$$

The last term, $\text{cov}[\bar{x}, \bar{y}]$, can be calculated with the observation that if x and y are chosen independently within the finite population, then $\text{cov}(x, y)$ in (A.3) will be zero in every state. Since in this case $G = G^*$, we have

$$\text{var}[\bar{x}] = \text{cov}[\bar{x}, \bar{y}] = G^*\text{E}[\bar{x}](1 - \text{E}[\bar{x}]) \qquad (A.4)$$

and we get an interesting formula for the variance of the average allele frequency, $\bar{x}$, in a finite population. Putting Eq. (A.4) into Eq. (A.3), we get Eq. (2.4).

References

Charnov, E.L., 1977. An elementary treatment of the genetical theory of kin selection. J. Theor. Biol. 66, 541–550.

Grafen, A., 1985. A geometric view of relatedness. Oxford Surv. Evol. Biol. 2, 28–89.

Grafen, A., 2006. Optimization of inclusive fitness. J. Theor. Biol. 238, 541–563.

Grafen, A., Archetti, M., 2007. Natural selection of social behaviour in viscous populations, in preparation.

Hamilton, W.D., 1964. The genetical evolution of social behaviour, I and II. J. Theor. Biol. 7, 1–52.

Lessard, S., 2005. Long-term stability from fixation probabilities in finite populations: new perspectives for ESS theory. Theor. Pop. Biol. 68, 19–27.

Lieberman, E., Hauert, C., Nowak, M.A., 2005. Evolutionary dynamics on graphs. Nature 433, 312–316.

Michod, R.E., Hamilton, W.D., 1980. Coefficients of relatedness in sociobiology. Nature 288, 694–697.

Moran, P.A.P., 1958. Random processes in genetics. Proc. Cambridge Philos. Soc. 54, 60–71.

Moran, P.A.P., 1962. Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.

Nowak, M.A., Sasaki, A., Taylor, C., Fudenberg, D., 2004. Emergence of cooperation and evolutionary stability in finite populations. Nature 428, 646–650.

Ohtsuki, H., Hauert, C., Lieberman, E., Nowak, M.A., 2006a. A simple rule for the evolution of cooperation on graphs. Nature 441, 502–505.

Ohtsuki, H., Nowak, M.A., 2006b. Evolutionary games on cycles. Proc. R. Soc. B 273, 2249–2256. DOI:10.1098/rspb.2006.3576.

Ohtsuki, H., Pacheco, J.M., Nowak, M.A., 2007. Evolutionary graph theory: breaking the symmetry between interaction and replacement. J. Theor. Biol. 246, 681–694.

Orzack, S.H., Hines, W.G.S., 2005. The evolution of strategy variation: will an ESS evolve? Evolution 59, 1183–1193.

Price, G.R., 1970. Selection and covariance. Nature 227, 520–521.

Proulx, S.R., Day, T., 2001. What can invasion analyses tell us about evolution under stochasticity in finite populations? Selection 2, 1–14.

Queller, D.C., 1985. Kinship, reciprocity and synergism in the evolution of social behaviour: a synthetic model. Nature 318, 366–367.

Ross, S.M., 1998. A First Course in Probability, Fifth ed. Prentice-Hall Inc, Englewood Cliffs, NJ.

Rousset, F., 2002. Inbreeding and relatedness coefficients: what do they measure? Heredity 88, 371–380.

Rousset, F., Billiard, S., 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. J. Evol. Biol. 13, 814–825.

Seger, J., 1981. Kinship and covariance. J. Theor. Biol. 91, 191–213.

Taylor, P.D., 1992a. Altruism in viscous populations—an inclusive fitness model. Evol. Ecol. 6, 352–356.

Taylor, P.D., 1992b. Inclusive fitness in a homogeneous environment. Proc. R. Soc. London B 249, 299–302.

Taylor, P.D., Frank, S., 1996. How to make a kin selection argument. J. Theor. Biol. 180, 27–37.

Taylor, P.D., Irwin, A., Day, T., 2000. Inclusive fitness in finite deme-structured and stepping-stone populations. Selection 1, 83–93.

Taylor, P.D., Day, T., Wild, G., 2007. Evolution of cooperation in a finite homogeneous graph. Nature 447, 469–472.

Wild, G., Taylor, P.D., 2004. Fitness and evolutionary stability in game theoretic models of finite populations. Proc. R. Soc. 271, 2345–2349.

Wild, G., Traulsen, A., 2007. The different limits of weak selection and the evolutionary dynamics of finite populations. J. Theor. Biol., in press.

Wilson, D.S., Pollock, G.B., Dugatkin, L.A., 1992. Can altruism evolve in purely viscous populations? Evol. Ecol. 6, 331–341.

Wright, S., 1943. Isolation by distance. Genetics 28, 114–138.