The infinitesimal model: Definition, derivation, and implications

CrossMark

N.H. Barton [a], A.M. Etheridge [b], A. Véber [c],*

[a] Institute of Science and Technology, Am Campus I, A-3400 Klosterneuberg, Austria
[b] Department of Statistics, University of Oxford, 24–29 St Giles, Oxford OX1 3LB, UK
[c] Centre de Mathématiques Appliquées, École Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France

ABSTRACT

Our focus here is on the *infinitesimal model*. In this model, one or several quantitative traits are described as the sum of a genetic and a non-genetic component, the first being distributed within families as a normal random variable centred at the average of the parental genetic components, and with a variance independent of the parental traits. Thus, the variance that segregates within families is not perturbed by selection, and can be predicted from the variance components. This does not necessarily imply that the trait distribution across the whole population should be Gaussian, and indeed selection or population structure may have a substantial effect on the overall trait distribution. One of our main aims is to identify some general conditions on the allelic effects for the infinitesimal model to be accurate. We first review the long history of the infinitesimal model in quantitative genetics. Then we formulate the model at the phenotypic level in terms of individual trait values and relationships between individuals, but including different evolutionary processes: genetic drift, recombination, selection, mutation, population structure, ... We give a range of examples of its application to evolutionary questions related to stabilising selection, assortative mating, effective population size and response to selection, habitat preference and speciation. We provide a mathematical justification of the model as the limit as the number M of underlying loci tends to infinity of a model with Mendelian inheritance, mutation and environmental noise, when the genetic component of the trait is purely additive. We also show how the model generalises to include epistatic effects. We prove in particular that, within each family, the genetic components of the individual trait values in the current generation are indeed normally distributed with a variance independent of ancestral traits, up to an error of order $1/\sqrt{M}$. Simulations suggest that in some cases the convergence may be as fast as $1/M$.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The infinitesimal model is a simple and robust model for the inheritance of quantitative traits, in which these are the sum of a genetic and a non-genetic (environmental) component, and the genetic component of offspring traits follows a normal distribution around the average of the parents; this distribution has a variance that is independent of the parental trait values, and, in a large outcrossing population, the variance remains constant despite selection. With inbreeding, the variance decreases in proportion to relatedness. Of course, selection may cause the distribution across the whole population to deviate from normality. The crucial point is that under the infinitesimal model, the distribution of genetic components *within families* remains normal, with variance that evolves in a way that is entirely determined by relatedness.

This model has its roots in the observations of Galton (1877, 1885, 1889), and their analysis by Pearson (1896, 1897). Fisher (1918) showed that trait values and their (co)variances can be broken down into components, and that the phenotypic observation of constant within-family variance is consistent with a large number of Mendelian factors, with additive effects. The limiting infinitesimal model can be extended to include all the main evolutionary processes: recombination, mutation, random sampling drift, migration and selection. The model is hardly new, yet there seems to be no agreement on what precisely is meant by the infinitesimal model, nor on the conditions under which it is expected to apply. Moreover, although it has long been central to practical breeding, where it forms the genetic basis for the *animal model*, it is relatively little used in evolutionary modelling (see Kruuk, 2004; Hill and Kirkpatrick, 2010 for a review).

This paper provides a summary of the model, together with a rigorous derivation, including control over its accuracy as an approximation. We show that its predictions about within-family variance can be accurate even with epistasis. The reason can be

understood intuitively, as follows. The classical theory of quantitative genetics gives a remarkably general description of evolution, in which the covariance in the values of a trait across individuals is a linear combination of a set of variance components, with coefficients determined by the probability of identity of sets of genes. Selection rapidly changes the trait mean, at a rate proportional to the additive genetic variance. However, when the trait depends on large numbers of genes, each of which makes a small contribution, selection has a negligible effect on the variance contributed by any individual locus. At the individual level, conditioning on the trait value hardly alters the distribution of effects of any one gene, at least in the short term; therefore, this distribution can be assumed constant. Importantly, it is not that allele frequencies do not change under the infinitesimal model: allele frequencies may change substantially due to random drift, mutation and migration; the key assumption is that selection only slightly perturbs the neutral distribution at any particular locus (Fisher, 1918; Robertson, 1960; Kimura, 1983, Ch. 6).

Our results here incorporate not only selection, but also mutation, random drift, population structure and some forms of epistasis. Dominance is left to future work. The evolutionary forces at work are captured by the actual pedigree of the population. Indeed, selection and structure pick out a particular pedigree, biased according to the trait values and the possible interactions between individuals. Thus, by conditioning on this pedigree and on the trait values in all previous generations, we are able to capture arbitrary forms of selection and population structure. The distribution of traits within families in the population is a multivariate normal distribution in which covariance is determined entirely by the pedigree and is independent of ancestral trait values. If some part of the pedigree or ancestral traits is unknown, then averaging with respect to the expected ancestral distribution, this multivariate normality is preserved. For example, it follows directly that conditioning on knowing just some of the trait values in the pedigree shifts the mean trait values in other families by a linear function of the conditioned values, but leaves variances within families unaltered.

After outlining the history of the infinitesimal model, we define it directly as a model for the distribution of phenotypes in a population; such a formal definition seems to be new. Initially, we implicitly assume an additive trait, but include all the usual evolutionary processes. For simplicity, we neglect linkage throughout. Having explained the phenotypic model, not only defining it at the level of the individual, but also showing how it can be simulated at the level of the population, we outline some of its applications. We then show that we can derive this infinitesimal model as the limit of a model of Mendelian inheritance, showing the conditions under which it is accurate and obtaining explicit bounds on the error. Finally, we show how the infinitesimal model extends to allow for epistasis, before presenting simulations that illustrate the main results.

We emphasise that our derivation of the infinitesimal model is distinct from earlier work, which used multi-locus models to analyse the effects of selection on complex traits (e.g. Bürger, 2000; Turelli and Barton, 1994; Kirkpatrick et al., 2002). The aim there was to connect population with quantitative genetics, and specifically, to find ways to approximate the effects of selection on the genetic variance, given a finite number of loci. In particular, Turelli and Barton (1994) investigated whether the trait distribution *across the whole population* could be approximated by a normal distribution. In contrast, here we aim to show that in the infinitesimal limit, the trait distribution *within families* is normally distributed, with a variance that is determined by the variance in the ancestral population and the pedigree relating individuals in those families, without making any detailed assumptions about the genetic basis of the trait, or about the form of the distribution of the trait across the population. Thus, we aim to recover the radical simplicity of quantitative genetics.

2. The classical model

2.1. History

Although the infinitesimal model is named for its justification as the limit of infinitely many Mendelian genes, it can be defined purely phenotypically, and its origins trace back well before the rediscovery of Mendel's work in 1900. Here, we summarise the origins of the infinitesimal model, after which we will formulate a precise definition at the phenotypic level, with no explicit genetic assumptions.

In one of the earliest quantitative discussions of heredity, Fleeming Jenkin (1867) argued that blending inheritance could have no effect in the long term: a white man stranded on an inhabited tropical island would leave offspring who, over successive generations, would approach ever closer to the dark-skinned native population. Davis (1871) pointed out that in a large and stable population, an individual is expected to leave two children, four grandchildren, and so on, so that his total expected contribution is constant through time. Nevertheless, if offspring are precisely intermediate between their parents, the range of variation in the population must necessarily decrease. Darwin saw this as a serious problem for his theory, which required a source of variation to counter blending inheritance. (See Bulmer, 2004, for a detailed discussion of Jenkin's argument.)

Francis Galton gathered extensive data on the inheritance of continuous traits, and introduced many ideas that are now central to quantitative genetics. In experiments with sweet peas, he showed that seeds of offspring grown from seeds of different weights followed a normal distribution with a mean that reverted towards the population mean, and with variance independent of the parents' weight: "I was certainly astonished to find the family variability of the produce of the little seeds to be equal to that of the big ones, but so it was, and I thankfully accept the fact, for if it had been otherwise, I cannot imagine, from theoretical considerations, how the problem could be solved" (Galton, 1877, p. 513). (In Galton's experiments with sweet peas, plants were self-fertilised, so that the variance in families is, in fact, expected to decrease.) He saw a similar pattern for human height, and showed that the joint distribution of offspring and mid-parent is bivariate normal (Galton, 1885). Moreover, he understood that the variance of the population could remain stable under the joint influence of random mating, reversion of offspring towards the population mean, and generation of variance amongst offspring. Galton (1877) calculated the equilibrium variance, allowing for Gaussian stabilising selection, a calculation next made by Bulmer (1971) and Cavalli-Sforza and Bodmer (1971), nearly a century later.

Galton (1885, 1889) tried to explain his observations by formulating his 'law of ancestral heredity', which divided an individual's phenotype into geometrically declining contributions from parents, grandparents, great-grandparents, …; he interpreted this contribution from distant ancestors as being due to inherited factors which have some probability, p, of being expressed in each generation. Bulmer (1998) shows that Galton's law is equivalent to the quantitative genetics of an additive trait, with p being replaced by the heritability, $h^2 = V_A/V_P$ (where V_P is the total phenotypic variance and V_A the additive genetic variance of the trait); however, h^2 may vary from trait to trait, whereas Galton assumed that it is a constant parameter of the mechanism of inheritance. Galton's model explains reversion of offspring towards the population mean as being due to expression of factors inherited from earlier generations (Lush, 1937, p. 47). In contrast, under Mendelian inheritance, reversion to the mean arises because selection acts on the phenotypic variance, V_P, whereas only additive genetic variation, V_A, is passed on; the deviation of offspring is therefore $h^2 = V_A/V_P$

times that of the selected parents. Pearson (1896, 1897) introduced matrix algebra and multiple regression, to put Galton's ancestral law on a firm mathematical basis. However, he treated the problem as the statistical description of a population of relatives, rather than following Galton in devising a mechanistic explanation of heredity (Magnello, 1998).

After 1900, there was a bitter dispute between those studying the Mendelian inheritance of discrete characters, and those biometricians who studied the inheritance of continuous traits (Provine, 1971). Pearson (1904a, 1904b, 1909) understood that Mendelian factors could account for continuous variation, but found that if there were complete dominance, correlations between relatives did not agree with observations (see Magnello, 1998). Yule (1902, 1907) showed that if incomplete dominance and random 'environmental' variation are included, then arbitrary correlations could be explained. However, these ideas were not developed further until Fisher's definitive (1918) paper.

During the following years, quantitative genetics developed quite separately from the population genetics of discrete genes. Fisher and Wright established the basic theory for correlation between relatives and for the effects of inbreeding, Wright was involved in practical animal breeding, and Haldane (1931) showed how selection on a trait affects the constituent alleles. However, the bulk of their work was on the evolution of single loci, and even the basic theory for the response of continuous traits to selection developed slowly. The change over one generation is implicit in Galton's regression of offspring on mid-parent, and the multivariate version is given by Pearson (1896). However, the classic 'breeders' equation' was only written in its modern form by Lush (1937); see Hill and Kirkpatrick (2010). Fisher's (1918) analysis of genetic variance was developed into a sophisticated theory in the 1950's (e.g. Cotterman, 1940; Henderson, 1953; Cockerham, 1954; Kempthorne, 1954; see Hill, 2014), but this did not become widely known. Quantitative genetics came back into contact with evolutionary questions through Robertson (1966), who formulated the 'secondary theorem', which states that the rate of change of a trait's mean due to selection equals its covariance with relative fitness. Robertson (1960) also showed that under the infinitesimal model, the ultimate cumulative response to selection equals the response in the first generation, multiplied by twice the effective population size; he showed that this can be understood through the increase in fixation probability of individual alleles caused by weak selection ($Ns \ll 1$). (We discuss this in more detail when we discuss applications of the infinitesimal model in Section 2.3.) Bulmer (1971) and Lande (1975) investigated the effect of stabilising selection and mutation on trait variance. It is striking that though these methods trace back to Galton and Pearson, they did not become widely used in evolutionary biology for more than 70 years. Indeed, the sophisticated 'animal model', widely used in animal breeding, has only been applied to analyse natural populations over the past 15 years (Kruuk, 2004).

Despite the revival of interest (both theoretical and empirical) in 'evolutionary quantitative genetics' in recent decades, the infinitesimal model itself has received little attention. Indeed, its origins are lost in the mists of time (M. Bulmer, W.G. Hill, pers. comm.). Bulmer (1971) showed that assuming a large number of unlinked loci with additive effects, the joint distribution of a set of relatives is multivariate normal, conditional on the parents; Lange (1978) gave a more detailed derivation. His aim was to find general mathematical conditions under which a polygenic model, based on a large number of loci, each having a small additive impact on a trait, implies a multivariate normal distribution for trait values of individuals in a group. Assuming either no inbreeding or no dominance variance, he provides conditions for a central limit theorem for polygenic-trait values in a pedigree. He assumes that all loci are in linkage equilibrium, that there is no assortative mating or

epistasis and that the number of chromosomes goes to infinity. In contrast to our work here, he considers only one generation of reproduction and he does not control the rate of convergence, nor the impact of conditioning on trait values of parents. Again assuming additivity, Dawson (1997) showed that certain kinds of linkage disequilibrium could cause the distribution amongst offspring to depend on the parental values. Bulmer (1974) and Santiago (1998) extended the infinitesimal model to allow for linkage. Appendix B of Turelli and Barton (1994) gave a general treatment of epistasis, which allows for linkage and multiple alleles. They showed that provided that kth order epistatic coefficients scale correctly with the number of loci, M, then the effect of selection on the trait depends only on the variance of effects at each locus, and the linkage disequilibria between them. The additive genetic variance will change slowly under selection, and can be assumed constant for $o(\sqrt{M})$ generations. However, their treatment did not include mutation, population structure, or random drift.

2.2. Definition of the phenotypic model

We begin by defining the infinitesimal model in terms of the phenotypic distribution. In Section 3.1, we derive it as the limit of a large number of Mendelian alleles with additive effects, and that underlying additivity will be implicit in our discussion in this section. However, in Section 3.2 we show that under some conditions the model can be extended to include epistasis and the phenotypic model will, just as in the classical case which we now describe, be determined by systems of recursions for the segregation variance between siblings.

For simplicity, in this section, we ignore non-genetic contributions to the trait; an environmental noise will be explicitly incorporated in our derivations in Section 3.1. We also consider a single trait, but it is straightforward to extend the analysis to multiple traits.

2.2.1. The basic model
Consider first the simplest case, of a purely additive trait in a large outcrossing population. Then, the infinitesimal model states that offspring follow a Gaussian distribution around the mean of the two parents, with a variance V_0, that is constant, independent of the parents' values. With random mating, the population as a whole rapidly converges to a Gaussian with variance $2V_0$. To see this, note that if the variance in the parental population is V_1, then that of the mean of two parents sampled at random is $V_1/2$, and so that of the offspring generation is $V_1/2 + V_0$: at equilibrium, $V_1 = 2V_0$; that is half the variance is between families, and half within them. Selection can readily generate arbitrary non-Gaussian distributions: for the population as a whole, we are free to choose any distribution of phenotypes (but within families the distribution remains Gaussian). However, in the absence of selection such a distribution rapidly relaxes back to a Gaussian with variance $2V_0$; the kth order cumulants decay by a factor 2^{1-k} per generation, for $k \geq 3$ (Bulmer, 1980). This is illustrated in Fig. 1.

2.2.2. Haploids versus diploids
In this simplest case, it makes no difference whether we follow haploids or diploids. However, the distinction becomes evident when we consider inbreeding and random drift. We can choose to follow haploid individuals, which mate to produce diploids that immediately undergo meiosis to produce the next haploid generation. Alternatively, we can follow diploid individuals, which produce haploid gametes via meiosis, and then immediately fuse to produce the next diploid generation. This results in two distinct approaches to modelling, both of which we describe below.

With no selection, whether we track haploids or diploids makes no fundamental difference. However, when we select, we condition the individual's full genotype on the value of a polygenic trait;
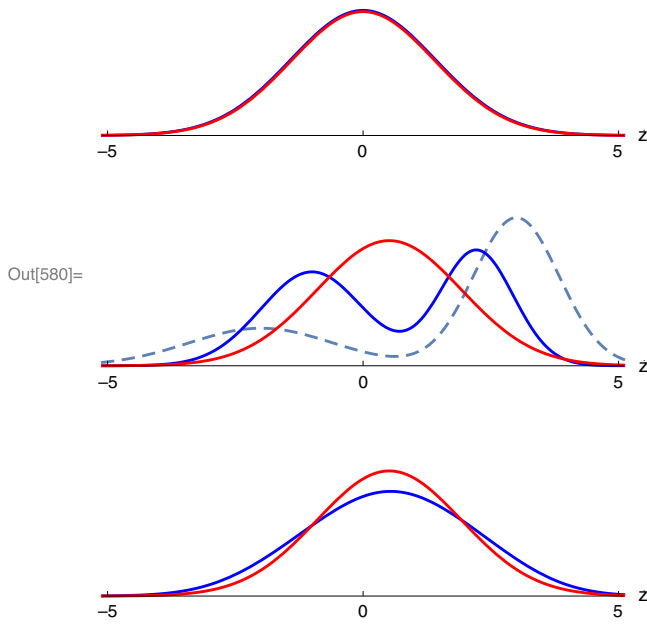
Fig. 1. Under the infinitesimal model, changes in the shape of the trait distribution across the population are due to linkage disequilibrium. Top: A large random-mating population at equilibrium, in the absence of selection. The genetic component of trait values follow a normal distribution, with variance divided equally within and between families. Middle: After disruptive viability selection (fitness shown by the dashed line), followed by random mating, the distribution of trait values is bimodal (blue), with increased mean. The underlying genic component of the distribution (red), which would be reached at linkage equilibrium, has the same (increased) mean, but a variance unchanged from the initial value. Bottom: After a further round of random mating, with no selection, allele frequencies, and hence the genic component (red) are the same. The distribution of trait values (blue) has relaxed towards this underlying normal distribution, as linkage disequilibria decrease as a result of segregation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

it is then clearly important whether we measure the trait at the haploid or the diploid stage. In principle, selection could act at both stages, but we do not consider this complication. For simplicity, in our derivation in Section 3 we concentrate on the haploid case.

2.2.3. Identity by descent and the segregation variance

In this section, we show how to incorporate inbreeding into the infinitesimal model. We explain this in modern terms, referring to genes, but emphasise that the formal definition of the infinitesimal model does not require that discrete genes be specified — only inbreeding coefficients, which can be calculated from the pedigree, are required.

As before, the mean trait value in offspring is the midpoint of the parents' trait values. Variation between siblings is generated by the random segregation of genes from the parental genotypes. To the extent that the genomes that come together in meiosis are related, this segregation variance will be reduced in proportion to the fraction of genes that they share. Imagine an ancestral population, whose members are all unrelated. We suppose that after one round of reproduction all families segregate with variance V_0. The current population descends from this reference population via an arbitrary pedigree. The relation between haploid individuals i, j is described by $F_{i,j}$, the probability that homologous genes descend from the same ancestral gene — that is, are identical by descent from the reference population. Since we are ignoring linkage, and the trait is additive, the variance amongst the haploid offspring from haploid parents i, j, is just $V_0(1 - F_{i,j})$.

For diploids, $F_{i,j}$ is defined to be the probability of identity between two genes, one from i, and one from j; when $i = j$, $F_{i,i}$ is

defined to be the probability of identity by descent of two *distinct* genes in the diploid individual i. Meiosis in i generates segregation variance proportional to $(1 - F_{i,i})$. The value of an additive trait in a diploid is the sum of equal contributions from each haploid gamete, and so the segregation variance is $V_0\left(1 - (F_{i,i} + F_{j,j})/2\right)$. To see this, one can note that segregation occurs independently to create the two parental gametes and, for each of them, conditional on not being identical by descent, the ancestral genes are two independent samples from the initial population with variance V_0. This yields an expression for the segregation variance of the form

$$\frac{1}{4}(1 - F_{i,i})2V_0 + \frac{1}{4}(1 - F_{j,j})2V_0 = V_0\left(1 - \frac{(F_{i,i} + F_{j,j})}{2}\right).$$

We have defined the infinitesimal model in terms of a constant genetic variance, V_0, in a reference population, together with a matrix of identity by descent. The entries in the matrix increase over time, until they are all one, and all variation is lost. However, we could instead follow the matrix of segregation variance between pairs of individuals. This process evolves with time, but it is Markovian (i.e., depends only on the current state), and it has the advantage that it does not require that we define an ancestral reference population at some arbitrary time in the past. As we shall see below, when we derive the infinitesimal model as a limit of a model of Mendelian inheritance, it is also convenient when we introduce mutation. For haploids, we define $C_{i,j}$ as the variance amongst offspring from two haploid parents, $C_{i,j} = V_0\left(1 - F_{i,j}\right)$. For diploids, as we saw above, the variance between siblings depends only on the identity between distinct genes in the parents, and not on the relationship between the two diploid parents. We define $C_{i,j} = V_0\left(1 - F_{i,j}\right)$, just as for haploids, but now, the variance amongst the diploid offspring of diploid individuals i, j is $\left(C_{i,i} + C_{j,j}\right)/2$.

Although $F_{i,j}$ is defined through the probability of identity by descent of discrete genes, it can (in principle) be measured purely through phenotypic measurements of the variance amongst offspring; this is perhaps clearer if we work with the $C_{i,j}$. The classical infinitesimal model is based on the assumption that the $C_{i,j}$ depend only on the pedigree relationship between i and j, and are independent of which trait is measured (to within the constant factor V_0), and (given the pedigree) of the trait values of the parents. In general, we think of a trait as being the sum of a genotypic value and an environmental deviation, independent of the underlying genetic values. We shall explicitly incorporate this non-genetic variation when we derive the infinitesimal model as a limit of Mendelian inheritance in Section 3. For the moment, we assume additivity and ignore environmental variation, so that the trait value is equal to the genotypic value, which in turn equals the *breeding value*. The breeding value of an individual is defined to be twice the mean deviation from the average phenotypic value, when it is crossed with a randomly chosen individual. The genotypic value can, in principle, be measured as the mean of large numbers of cloned individuals, and the breeding value can be measured through the mean of offspring from crosses made with randomly chosen mates. So, the infinitesimal model can be defined without identifying any specific genes.

2.2.4. Recursions for identity by descent

In a randomly mating population of N haploid individuals, reproducing under the Wright–Fisher model, the expected identity is $F = 1 - (1 - 1/N)^t$ after t generations. However, we consider the general case, where $F_{i,j}$ may vary arbitrarily between pairs. For haploids, $F_{i,i} = 1$ by definition. The recursion for F can be written in terms of a *pedigree matrix*, $P_{i,k}(t)$, which gives the probability that a gene in i in generation t came from parent k in the generation $(t - 1)$; each row has two non-zero entries each with value $1/2$,

unless the individual is produced by selfing, in which case there is a single entry with value 1. Thus,

$$F_{i,j}(t) = \sum_{k,l} P_{i,k}(t)P_{j,l}(t)F_{k,l}(t-1) \quad (i \neq j), \quad F_{i,i}(t) = 1. \tag{1}$$

For diploids, the corresponding recursion for F is

$$F_{i,j}(t) = \sum_{k,l} P_{i,k}(t)P_{j,l}(t)F_{k,l}^*(t-1), \tag{2}$$

where

$$F_{k,l}^* = F_{k,l} \quad \text{if } k \neq l, \qquad F_{k,k}^* = \frac{1}{2}\left(1 + F_{k,k}\right).$$

The quantity $F_{k,l}^*$ is the probability of identity of two genes drawn *independently* from k, l; if $k = l$, then the probability of drawing the same gene twice is one half.

If we work with the segregation variances, $C_{i,j}$, then the recursion for haploids is

$$C_{i,j}(t) = \sum_{k,l} P_{i,k}(t)P_{j,l}(t)C_{k,l}(t-1) \quad (i \neq j), \quad C_{i,i}(t) = 0, \tag{3}$$

and for diploids is

$$C_{i,j}(t) = \sum_{k,l} P_{i,k}(t)P_{j,l}(t)C_{k,l}^*(t-1),$$

where

$$C_{k,l}^* = C_{k,l} \quad \text{if } k \neq l, \qquad C_{k,k}^* = \frac{1}{2}C_{k,k}.$$

Note that the variance in the base population, V_0, does not appear explicitly: the future dynamics are entirely determined by the variation that is released through recombination between any pair of genomes. Although the precise recursions that we have written down are particular to the additive model, analogous recursions characterise the segregation variance in our more complex models that incorporate house of cards mutation and epistasis. The key fact will be that it is the pedigree relatedness between individuals that drives the recursions. As long as the variances in the parental population are sufficiently large relative to the effect of individual alleles, knowing the trait values of the parents has a negligible effect on the segregation variance; in other words, the infinitesimal model remains valid.

2.2.5. Simulating the infinitesimal

The infinitesimal model can be simulated either at the level of the individual, or the population. An individual-based simulation must follow the breeding values of each individual, z_i, and the relatedness between individuals, $F_{i,j}$. Extension to multiple traits would require that we follow vectors of breeding values. Since the main computational effort is in calculating the matrix of identities, this is not much extra burden. The matrix of identities can be iterated efficiently by representing Eqs. (1) and (2) in matrix form, but the size of the population is ultimately limited by the memory needed to store $F_{i,j}$. However, in large populations $F_{i,j}$ typically approaches the same small value between almost all pairs; thus, it can be approximated as a constant plus a sparse matrix that tracks close relatives. Populations of many thousands can then be simulated (e.g. Barton and Etheridge, 2011).

Provided that the pedigree determined by the matrix $F_{i,j}$ is not too skewed towards large contributions from particular individuals, then we can also simulate very large populations by following the distribution of the trait and the *average* value of $F_{i,j}$ through time. To do this, first, the continuous trait distribution must be approximated by a discrete vector; selection on the trait is represented by multiplying the trait distribution by the fitness. Since reproduction involves a convolution between the parents'

distributions and the Gaussian distribution of offspring, it is convenient to follow the (discrete) Fourier transform: convolution of distributions corresponds to multiplication of their transforms (e.g. Polechova and Barton, 2005) In each generation, there must be a conversion between the distribution and its transform, which can be done efficiently using the fast Fourier transform algorithm (Gauss, 1866; Cooley and Tukey, 1965). Evidently, the approximation that all individuals are related by the same $F_{i,j}$ will not always be realistic, in which case an individual based approach becomes essential.

2.2.6. Mutation

Pragmatically, for traits determined by a very large number of loci, mutation can be included by scaling the recursion to account for alleles that are replaced by mutants and adding a constant, which may depend on t, to every element of the matrix $C_{i,j}$ in each generation to account for the variance introduced by mutation (Wray, 1990). Mutation may be biased: in particular, we expect mutation to decrease traits that have been under directional selection, and so to decrease fitness. This can be described by scaling the mean of the offspring, by a constant $(1 - \mu)$ say, and shifting it by another constant δ_μ. Under this extension to the infinitesimal model, μ is assumed constant and the distribution of offspring, conditional on their parents, is assumed Gaussian. Segregation variance is partitioned into portions attributed to the variance present in the base population and that attributed to mutations arising in successive generations. We shall see that we can obtain such a model by introducing 'house of cards' mutation to our model of an additive trait determined by a large number of Mendelian factors. The recursion for the segregation variance is then given by

$$C_{i,j}(t) \approx (1 - 2\mu) \sum_{k,l} P_{i,k}(t)P_{j,l}(t)C_{k,l}(t-1) + \mu V_m$$

(see Eq. (13)), where μV_m can be interpreted as the additive genetic variance introduced by new mutations in each generation. Here μ is the probability of a mutation at a given locus in a given individual in one generation of reproduction and so can be expected to be very small. In order for mutation to have an appreciable effect on the overall trait over a small number of generations, we must assume that V_m, the variance of the sum of the allelic effects of the new mutations arising in a given generation, is large. Here we see a tradeoff: the larger the allelic effect of mutations, the less accurate an approximation the infinitesimal model becomes.

2.2.7. Population structure and gene flow

When defined in terms of individual trait values and relationships, the infinitesimal model automatically incorporates arbitrary population structure. For example, a well-mixed reference population may split into separate demes, so that the probability of identity F between genes from within the same population would increase relative to that between populations. If the demic structure were permanent, then it would be more natural simply to follow the segregation variance $C_{i,j}$, which would be higher from crosses between populations than within them. If, for whatever reason, the sub-populations in different demes diverge, then the distribution of trait values within a single deme will not be Gaussian, even though, under the infinitesimal model, the distribution amongst offspring of given parents will always be Gaussian. The same is true for any pedigree which is not 'well-mixed'; the key point is that parental trait values determine the mean trait value among offspring, but the segregation variances that determine the variance of offspring traits within each family are completely determined by the pedigree, which can reflect arbitrary mating patterns, family structure, and spatial subdivision.

The power of the infinitesimal model in capturing population structure comes at a price; for a large population it may not be

practicable to trace the breeding values and relatedness of all individuals. In that case, one could try to approximate the infinitesimal model, for example by assuming that the trait distribution within populations is approximately Gaussian, or that relationships (or equivalently, segregation variance) between individuals within subpopulations are homogeneous. Such approximations may be delicate, since they might need to take account of the reduction in effective migration rate, and the increase in the rate of random drift, due to selection. However, it is important to realise that the infinitesimal model itself can be defined at the level of individuals, even when trait distributions across the whole population are far from Gaussian, and relationships are heterogeneous.

2.2.8. Multivariate normality

Even though the trait distribution across the whole population may be far from Gaussian, the multivariate normal will play a central rôle in our analysis. The deviations of individual trait values from their mid-parent value will be described by a multivariate normal whose variance–covariance matrix is independent of the parental traits. We establish this result conditional on knowing the pedigree and all ancestral traits within that pedigree, but equally we could have conditioned on the values of any subset of relatives and the same would hold true: expected trait values would be a linear functional (determined by the pedigree) of the values on which we conditioned, but segregation variances would be unchanged.

Bulmer (1971) and Lange (1978) showed that the unconditional joint distribution of traits converges to a multivariate normal. Lange also allows for some linkage among loci by allowing inheritance to be dependent among loci at distance at most q from one another. We do not include linkage in our analysis, although as long as recombination is sufficiently fast, our results should hold true, but whereas the rate of convergence to the infinitesimal model when we have unlinked loci is $1/\sqrt{M}$, with linkage it will be $1/\sqrt{M^*}$, where M^* is an 'effective' number of loci. Bulmer assumed random mating, while Lange's proof is for individuals related through a given pedigree. However, as Lange remarks, his result gives no control of the rate of convergence. This is essential if we wish to approximate the conditional distribution, knowing some ancestral trait values. It is also needed in assessing the accuracy of the infinitesimal model as an approximation, and is the focus of our derivation.

2.2.9. Epistasis

Thus far, we have defined the infinitesimal model for the additive case. Evidently we cannot extend it to arbitrary epistatic interactions. For example, if Z is a purely additive trait, then Z^2 is a sum of additive and pairwise epistatic components. Since the square of a normally distributed random variable is not normally distributed, the infinitesimal model must clearly break down. However, under some conditions (which we lay out in Section 3.2), even though there can be significant variance due to epistatic components, and the mean trait value among offspring in a family will no longer be the average of the parental traits, the key prediction of the infinitesimal model still holds with epistasis: the variance of the trait distribution within a family will depend only on the pedigree and the variance in the ancestral population.

In general, with epistasis the individual phenotype is

$$z = z_0 + \sum_U \eta_U + E, \tag{4}$$

where the sum is over the average effects η_U of all sets U of distinct loci, and E is a random non-genetic component that is assumed to have a distribution independent of genotype, and independent between individuals (see e.g. Chapter 7 in Lynch and Walsh, 1997).

The sets of genes U descend from a homologous set in the base population, which in general will be scattered over many individuals. The η_U are defined as the marginal effects of the set of genes U, that remain after accounting for the effects of all subsets of U. If the base population is in linkage equilibrium, then the η_U are uncorrelated. The sum of the variances contributed by sets of size $|U| = k$ is the k th order epistatic variance, $\sum_{|U|=k} V_U = V_{A(k)}$. In contrast to the additive case, we see correlations between the deviations ΔZ from the mid-parental trait values of distinct individuals. The covariance between two distinct individuals is

$$\mathrm{cov}(\Delta Z, \Delta Z') = \sum_{|U| \geq 2} V_U F_U, \tag{5}$$

where F_U is the probability that the set of genes U in the two individuals are all identical by descent. If loci are unlinked, then this depends only on the number of genes in the set, so that $F_U = F_k$ where $k = |U|$; F_k is given by a recursion on the pedigree similar to that described above for pairwise identities (corresponding to F_1). It is more complicated, because we need to track the probability of identity for genes in up to $2k$ individuals. However, if identity at different loci is uncorrelated, then $F_k = F_1^k$, where F_1 is the pairwise recursion defined above. Unless inbreeding is intense, this is typically a good approximation (Barton and Turelli, 2004).

This partition of genetic variance into components applies regardless of the number of loci. Crucially, however, if the joint distribution of the components of trait values across the pedigree is multivariate normal, then the mean and covariance completely define that distribution. In the following, we outline the proof that the distribution of trait values is indeed multivariate normal in the case when we allow just pairwise epistatic interactions, provided that the total allelic effect of any particular gene is not too large, and indicate how this could be extended to also include higher order interactions.

We emphasise that the components of phenotype, η_U, and the corresponding variances, V_U, are defined relative to the base population. In any particular descendant population, the trait mean will differ as a result of mutation (which we exclude from our analysis in the epistatic case), selection and random drift. With epistasis, the effects relative to the new population will be different, and so the variance components defined for this descendant population will also differ. This can lead to the 'conversion' of epistatic variance into additive variance (Barton and Turelli, 2004). We do not consider this issue here, since we always define variance components relative to the base population. However, it is straightforward to change the reference point.

2.3. Applications of the infinitesimal model

We have defined the infinitesimal model in terms of individual trait values and relationships between individuals, without referring explicitly to discrete genes. This is essentially the 'animal model', which is the basis for practical animal breeding, though extended to include mutation. In practical applications, the 'animal model' is typically applied to a given pedigree, and is used to estimate breeding values and genetic variances conditional on that pedigree (Hill, 2014). In recent years, it has also been applied to parameter estimation in natural populations (Kruuk, 2004). However, it has been surprisingly little used for addressing evolutionary questions. Here, we illustrate the power of the infinitesimal model as a tool for understanding aspects of evolution, by presenting a range of examples related to stabilising selection, assortative mating, effective population size and response to selection, habitat preference and speciation.

Perhaps the simplest non-trivial application of the infinitesimal model is to understand stabilising selection (Galton, 1877; Slatkin, 1970; Bulmer, 1971; Cavalli-Sforza and Feldman, 1976;

Lande, 1975). Suppose that the distribution of a trait in a parental population is Gaussian with mean $\bar{z}$ and variance V_g. If fitness is a Gaussian function of trait value, with mean z_0 and variance V_s, then after selection the new trait distribution is obtained by multiplying the density of a trait by the fitness of that trait and renormalising to have total mass one, resulting in a Gaussian with mean

$$\bar{z} + (z_0 - \bar{z}) \frac{V_g}{V_g + V_s},$$

and variance

$$\frac{V_s V_g}{V_s + V_g}.$$

After random mating and reproduction under the infinitesimal model (without mutation), the mean remains the same, and the variance is

$$\left(V_0 + \frac{V_s V_g}{2 \left(V_s + V_g \right)} \right),$$

where V_0 is the segregation variance, and $\frac{V_s V_g}{2(V_s + V_g)}$ is the variance of the mean of two randomly mated parents. Therefore, at equilibrium, the genetic variance across the population immediately after reproduction is

$$V_g = \frac{V_s}{4} \left(2 \left(\frac{V_0}{V_s} \right) + \sqrt{1 + 12 \left(\frac{V_0}{V_s} \right) + 4 \left(\frac{V_0}{V_s} \right)^2} - 1 \right), \qquad (6)$$

which decreases from $2V_0$ (the value we obtained for a neutral population) when stabilising selection is weak ($V_s \gg V_0$), to V_0 when stabilising selection is so strong as to eliminate all variation ($V_s \ll V_0$).

The effect of assortative mating is more surprising. Suppose that the relative contribution to the next generation of pairs of individuals with trait values z_1, z_2 is proportional to

$$\exp \left(-\frac{(z_1 - z_2)^2}{2V_a} + \frac{(z_1^2 + z_2^2)}{2\omega} \right),$$

where ω is chosen so that there is no direct selection on individuals (i.e., there is no marginal effect of the trait on individual fitness; see, for example, Appendix 4 of Polechova and Barton, 2005 for an expression for the ω that achieves this). The mean does not change, but assortment results in a higher variance in the mid-parent value than under random mating. (Fisher, 1918; Crow and Felsenstein, 1968). To understand this, note that because the contribution of pairs of individuals is greater for individuals with similar trait values, more extreme traits are less likely to be pulled towards the mean; indeed in the most extreme case, individuals would only reproduce with others with an identical trait value and so the distribution of mid-parent values would have the same variance as that of the whole parental population. Provided that assortment is not too strong ($V_a > 4V_0$), there is an equilibrium genetic variance across the population $2V_0 (V_a - 2V_0) / (V_a - 4V_0)$. However, if assortment is very strong, ($V_a < 4V_0$), the variance increases without limit. In reality, either the infinitesimal model would break down as genetic limits to the trait are approached, or stabilising selection would prevent indefinite divergence.

This simple model makes the important point that assortative mating alone can lead to indefinite divergence, and, ultimately, speciation (Polechova and Barton, 2005). The infinitesimal model can also be extended to model the joint evolution of habitat preference and viability in two different niches, both being represented as continuous traits (Barton, 2010). Assortative mating, and eventual reproductive isolation, then arise as a by-product of preferences for different habitats, provided mating occurs within those habitats (Diehl and Bush, 1989).

In a random-mating population of effective size N_e of diploid individuals, the segregation variance decreases by a factor $(1 - 1/2N_e)$ per generation. The infinitesimal model predicts that the response to steady directional selection will decrease at the same rate and so, summing over generations (a geometric sum), Robertson (1960) found the total response to selection to be just $2N_e$ times the change in the first generation. He also found an alternative derivation of the same result (for $N_e s$ small), by considering the increase in fixation probability of neutral alleles. Hill (1982) extended Robertson's work to include mutation that introduces genetic variance at a rate V_m. The genetic variance then reaches an equilibrium between mutation and random drift, $V_g = 2N_e V_m$, and the response to directional selection is proportional to this variance. In large populations, selection will tend to reduce genetic variance that is due to alleles with large $N_e s$; however, some component of the genetic variance will be due to more weakly selected alleles with $N_e s$ small. In a survey of selection experiments, Weber and Diggins (1990) showed that the ratio of the response to selection after fifty generations to that after just one generation is somewhat less than predicted by the infinitesimal model. Zhang and Hill (2005, Fig. 6) showed that, at least for Drosophila experiments, this reduced response can be explained either by alleles of large effect or by linkage. Thus in these experiments the response to selection is largely explained by the infinitesimal model.

Selection on heritable traits can greatly inflate the rate of random drift: genes that find themselves in a fit genetic background in one generation will tend to be in a fitter background in subsequent generations, even if all loci are unlinked; this correlation in fitness across generations increases the rate of sampling drift (Robertson, 1961). The infinitesimal model can be used to estimate this inflation, by finding the variance in reproductive value (Barton and Etheridge, 2011), and the decrease in fixation probability of favourable alleles (Weissman and Barton, 2012).

Apart from these few examples, the infinitesimal model has hardly been used in evolutionary modelling. It should not be confused with two other models that have been used more extensively. Kimura (1965) investigated the distribution of effects of alleles at a *single* locus, and approximated this *continuum-of-alleles* model by a Gaussian; Lande (1975) developed this model to investigate maintenance of variation by mutation, despite stabilising selection. This is a quite different approach from the infinitesimal model, which requires no strong assumptions about the distribution of effects at each locus, and which does not assume a Gaussian distribution of trait values. A second model that bears a superficial resemblance to the infinitesimal model is the *hypergeometric* or *symmetric* approximation, which assumes that the trait is determined by additive loci of equal effect, and that all genotypes that give the same trait value are equally frequent (Kondrashov, 1984; Doebeli, 1996; Barton and Shpak, 2000). This is a very strong assumption; the symmetry between genotypes may hold under disruptive selection, but is unstable under stabilising selection, when any one of the many optimal genotypes tends to fix (Wright, 1935).

3. The infinitesimal model as the limit of Mendelian inheritance

In this section, we turn to a justification of the infinitesimal model as a limit of a model of Mendelian inheritance, when trait values are determined by a large number of Mendelian factors, each of small effect. By performing the analysis carefully, we determine the accuracy of the infinitesimal approximation as the number of loci tends to infinity. We start in the classical setting, in which traits are additive. Using the same arguments we also include 'house of cards' mutation, but we see a tradeoff: for the infinitesimal model to apply with mutation, there must be a large

number of mutant loci each of small effect, but the overall effect is then proportional to the probability of mutation per locus, per individual, per generation, which should typically be very small. We then indicate how the results can be extended to include some forms of epistasis. In all cases, observed trait values are assumed to be composed of a genetic component plus an independent environmental noise. It is the genetic component that follows the infinitesimal model, that is for which the distribution of trait values within families follows a multivariate normal distribution with a variance that depends on the pedigree, but not on the trait values of ancestors in the pedigree. The observed trait values will not follow the infinitesimal model, but in the special case where the noise is itself normally distributed, the observed traits within families will be asymptotically normally distributed as the number of loci tends to infinity, and we shall write down recursions for the mean vector and the variance–covariance matrices.

In all cases, the key issue is this: knowing the segregation variance V_0 in our base population and the pedigree F (or, equivalently, the matrices C of segregation variances in previous generations), how close is the segregation variance of the offspring of parents i and j to being independent of the trait values of those parents? It is important to note that this does *not* say that the *pedigree* is independent of the trait value; indeed, for a population undergoing artificial selection, for example, one can expect a strong dependence between trait values and pedigree. We also emphasise that trait values across the whole population can be very far from normal; it is the offspring within families that follow a multivariate normal distribution.

One necessarily expects some dependence of segregation variance on trait values: if the possible trait values are bounded, with a single genotype giving the largest value, say, then meiosis between two copies of this most extreme haploid type, or the products of meiosis from a diploid with the most extreme value, would have zero variance. For any trait values that are close to the extremes of what is possible, so that few genotypes produce these values, segregation variance will be radically reduced; the derivation of the infinitesimal model depends on there being a very large number of genotypes compatible with each trait value, so that conditioning on the trait does not give significant information about the underlying genotype frequencies.

In order to understand why for 'typical' trait values, knowing the trait value for an individual provides very little information about the allelic effect at a particular locus, it is instructive to consider a simple example. The argument we use is similar to that on p. 402 of Fisher (1918), where it is expressed in terms of a regression. Suppose that a particular trait is determined by the sum of allelic effects at M independent loci, with the allelic effect at the l th locus being $\eta_l/\sqrt{M} = \pm 1/\sqrt{M}$ with equal probability. Now suppose that we condition on the trait value being $k/\sqrt{M}$. For definiteness, we take M and k both to be even. What is the conditional probability that $\eta_1 = 1$? An application of Bayes' rule gives

$$\mathbb{P}\left[\eta_1 = 1 \,\middle|\, \sum_{l=1}^{M} \frac{\eta_l}{\sqrt{M}} = \frac{k}{\sqrt{M}}\right]$$

$$= \frac{\mathbb{P}\left[\sum_{l=1}^{M} \eta_l = k \,\middle|\, \eta_1 = 1\right]}{\mathbb{P}\left[\sum_{l=1}^{M} \eta_l = k\right]} \mathbb{P}[\eta_1 = 1]$$

$$= \frac{\mathbb{P}\left[\sum_{l=2}^{M} \eta_l = k - 1\right]}{\mathbb{P}\left[\sum_{l=1}^{M} \eta_l = k\right]} \mathbb{P}[\eta_1 = 1]$$

$$= \frac{\frac{1}{2^{M-1}} \binom{M-1}{(M+k-2)/2}}{\frac{1}{2^{M}} \binom{M}{(M+k)/2}} \mathbb{P}[\eta_1 = 1]$$

$$= \left(1 + \frac{k}{M}\right) \mathbb{P}[\eta_1 = 1].$$

For large M, a 'typical' value for k is $\mathcal{O}(\sqrt{M})$, and then this calculation says that, for any particular locus, the chance that it 'notices' the conditioning is $\mathcal{O}(\sqrt{M}/M) = \mathcal{O}(1/\sqrt{M})$. On the other hand, at the extremes of what is possible ($k = \pm M$) the value of the trait gives complete information about the allelic effect at each locus.

As we see below, essentially the same argument applies to much more general models for the allelic effects at each locus: for the infinitesimal model to be a good approximation, the observed parental trait values must not contain too much information about the allelic effect at any given locus, and for this to hold true, the parental traits must not be too extreme.

3.1. The additive case with mutation and environmental noise

Once we have established our notation, we shall set out the derivation in the strictly additive case. To simplify notation we omit mutation in this outline, only indicating its effect on the statement of the results. The details of the proofs are provided in the appendices, where we do incorporate (house of cards) mutation. We also suppose that the observed trait value is a combination of a genetically determined component and an environmental noise which, for simplicity, we take to be an independent draw from a mean zero Gaussian distribution for each individual in the population.

Laying out this argument carefully enables us to identify conditions under which our results can be modified to include epistasis, which we illustrate through a simple example. In this case, even for the genetic component of the trait, the mean value among offspring in a family will not simply be the average of the parental values; what we can prove is that the variance of trait values within the family is independent of the parental trait values and is determined by the segregation variance in the base population plus the pedigree.

Throughout we concentrate on the haploid case, although, at the expense of more complicated notation and formulae, the approach extends to the diploid case. Indeed, following Lange (1978) and Abney et al. (2000), we anticipate also being able to include dominance. A more detailed study of epistasis and dominance is deferred to later work.

The formulae that follow are, at first sight, a little daunting. To make them slightly easier to navigate, we impose some conventions in our notation. Table 1 summarises all our notation.

Assumptions and Notation

We reserve the indices i and j for individuals in our population, whereas l and m are used for loci, of which there are M. Generation number will be indexed by t (but will mostly be implicit). The total population size in generation t is N_t.

1. *Allelic effect at locus l.* We denote the allelic effect at locus l in the jth individual by $\eta_l^j/\sqrt{M}$. We centre η_l^j relative to the mean allelic effect at locus l in the ancestral population. The scaling of $1/\sqrt{M}$ ensures that the additive genetic variance is of order one. The random variable η_l^j is assumed to be uniformly bounded over all loci, with $|\eta_l^j| \leqslant B$. We sometimes refer to it as the *scaled* allelic effect.
2. *Genetic component of the trait value.* The genetic component of the trait value in the j th individual in the present generation will be denoted by Z^j. It will always be written as $\bar{z}_0$, its average value in the ancestral population, plus a sum over loci of allelic effects. That is, in the notation just defined, the genetic component of the trait of the j th individual is

$$Z^j = \bar{z}_0 + \sum_{l=1}^{M} \frac{1}{\sqrt{M}} \eta_l^j. \tag{7}$$

Table 1
Notation.

Phenotypic model	
F	Matrix of expected identity by descent, Eqs. (1), (2)
C	Matrix of segregation variances, Eq. (3)

Traits	
η_l (or η_l^j)	Scaled allelic effect at locus l (in individual j)
$\widehat{\eta}_l$	” in ancestral population
$\breve{\eta}_l$	” after mutation
$\bar{z}_0$	Mean trait value in ancestral population
$\bar{z}_\mu$	$\bar{z}_0 + \mathbb{E}[\sum_{l=1}^{M} \breve{\eta}_l / \sqrt{M}]$
$\widehat{\sigma}_M^2$	$\sum_{l=1}^{M} \mathrm{Var}(\widehat{\eta}_l)/M \to \widehat{\sigma}^2$
$\breve{\sigma}_M^2$	$\sum_{l=1}^{M} \mathrm{Var}(\breve{\eta}_l)/M \to \breve{\sigma}^2$
Z^j	Genetic component of trait in individual j
E^j	Environmental component of trait in individual j
$\widetilde{Z}^j = Z^j + E^j$	Observed trait of individual j
η_{lm}	Scaled epistatic effect of loci $\{l, m\}$
$Z_{A(2)}$	Total pairwise epistatic effect

Inheritance	
$j[1], j[2]$	Labels of first and second parents of individual j
X_l^j	Indicator random variable that individual j inherits locus l from 'first parent'
$\mathcal{M}_l^j$	Indicator random variable that there is a mutation at locus l in individual j
μ	Mutation probability per locus per generation

Conditioning	
$\mathcal{P}(t)$	The pedigree up to generation t
$\widetilde{Z}(t)$	Observed traits of all individuals up to generation t
R^j	$Z^j - \mu \bar{z}_\mu - (1-\mu)(Z^{j[1]} + Z^{j[2]})/2$ (in main text $\mu = 0$)
$\Sigma_t^M, \Sigma_t^{M,\mu}$	Variance–covariance matrix of R^j, Eqs. (11) and (24)
Σ_t, Σ_t^μ	Limit as $M \to \infty$ of $\Sigma_t^M, \Sigma_t^{M,\mu}$ respectively
$C(\sigma^2, |z|)$	c.f. Eq. (12)

3. *Environmental noise and observed trait value.* We suppose that the observed trait value is

$$\widetilde{Z}^j = Z^j + E^j,$$

where the E^j are independent normally distributed random variables with mean zero and variance σ_E^2. The assumption of normality is not compulsory here. It has the advantage that the observed trait value will also follow a normal distribution in the limit as $M \to \infty$, which we can therefore characterise easily by its mean and variance.

4. *Ancestral population.* Although it is not strictly necessary, we assume that in generation zero, the individuals that found the pedigree are unrelated. They are sampled from an ancestral population in which all loci are assumed to be in linkage equilibrium. The genetic component of the trait value in the j th individual in generation zero is written as

$$Z^j = \bar{z}_0 + \sum_{l=1}^{M} \frac{1}{\sqrt{M}} \widehat{\eta}_l^j, \tag{8}$$

where the $\widehat{\eta}_l^j$ are independent for different values of j, with the same distribution as $\widehat{\eta}_l$ where $\mathbb{E}[\widehat{\eta}_l] = 0$ for all l. The random variables $\widehat{\eta}_l$ are assumed to be independent but not necessarily identically distributed.
We shall write

$$\widehat{\sigma}_M^2 = \frac{1}{M} \sum_{l=1}^{M} \mathrm{Var}(\widehat{\eta}_l)$$

and assume that $\widehat{\sigma}_M^2$ converges to a finite limit $\widehat{\sigma}^2$ as $M \to \infty$.

5. *Parents.* To distinguish the parents of an individual we order them. The symbols [1] and [2] will refer to the first and second parents of an individual, respectively. Thus $\eta_l^{j[1]}$ is the scaled allelic effect at locus l in the first parent of the j th individual. Similarly, $Z^{j[1]}$ will denote the genetic component of the trait value of the first parent of individual j. Note that we allow selfing, in which case parents 1 and 2 are identical.

Centring allelic effects relative to their mean in the ancestral population may seem unnatural, since these are quantities that we could never hope to measure, but in fact it is simply a mathematical convenience: only the variances of these quantities appear in our results.

Inheritance

We use Bernoulli random variables to encode the parent from which a locus inherits its allelic type. Thus we write $X_l^j = 1$ if the allelic type at locus l in the j th individual is inherited from the 'first parent' of that individual; otherwise it is zero. In particular, under Mendelian inheritance, $\mathbb{P}[X_l^j = 1] = 1/2 = \mathbb{P}[X_l^j = 0]$.

Of course, really there are independent Bernoulli random variables capturing the inheritance in each generation, but, since we only discuss transitions one generation at a time, we suppress that in our notation.

Conditioning

We shall use $\mathcal{P}(t)$ to denote the pedigree relationships between all individuals up to and including generation t and $\widetilde{Z}(t)$ will denote

the observed traits of all individuals in the pedigree up to and including the t'th generation. We shall be deriving the distribution of trait values in generation t conditional on knowing $\mathcal{P}(t)$ and $\widetilde{Z}(t-1)$. We distinguish $\mathcal{P}$ from the matrix of identities F, because conditional on $\widetilde{Z}(t)$, it is no longer true that tracing back through the pedigree, an allele is equally likely to come from either parent; indeed proving that this is *almost* the case is a key part of our derivation.

We note that, since selection acts on the trait, both the effect of selection and that of population structure is captured by the pedigree. Moreover, although we obtain our result by conditioning on knowing both the pedigree and *all* the observed traits of individuals in that pedigree up to and including the parental generation, exactly the same argument shows that we can condition on a subset of the trait values in the pedigree up to time $(t-1)$ and still obtain a multivariate normal for the distribution of traits within families in generation t, where the mean values are shifted by a linear function of the conditioned values, but variances within families will be unaltered.

Main result in the additive case

Our first aim is to understand, in the additive case, the genetic component of the traits, conditional on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1)$. The classical infinitesimal model deals with the genetic component of the traits. There, the mean trait within a family will be the midpoint of the parental traits, with a variance that is independent of their trait values. In our setting, since we only condition on the *observed* traits $\widetilde{Z}(t-1) = \underline{z}$, the conditioned quantities $Z^{j[1]}$ and $Z^{j[2]}$ are still random and so we must be slightly careful in stating our result. We will show that conditional on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1) = \underline{z}$,

$$\left(R^j\right)_{j=1,\dots,N_t} := \left(Z^j - \frac{Z^{j[1]} + Z^{j[2]}}{2}\right)_{j=1,\dots,N_t} \tag{9}$$

converges (in distribution) to a multivariate normal random variable with mean zero and diagonal covariance matrix Σ_t, with j th diagonal entry $(\Sigma_t)_{jj}$ given by the segregation variance among offspring of the parents of individual j.

More precisely, we show that, writing Φ for the cumulative distribution function of a standard normal random variable,

$$\left| \mathbb{P}\left[\frac{Z^j - \frac{Z^{j[1]}+Z^{j[2]}}{2}}{\sqrt{(\Sigma_t^M)_{jj}}} \le y \middle| \mathcal{P}(t), \widetilde{Z}(t-1) = \underline{z} \right] - \Phi(y) \right|$$
$$\le \frac{t}{\sqrt{M}} C\left(\underline{\Sigma}_t^M, \bar{\Delta}_t^M(\underline{z})\right), \tag{10}$$

where

$$(\Sigma_t^M)_{jj} = \frac{1}{4M} \sum_{l=1}^{M} \mathbb{E}\left[\left(\eta_l^{j[1]} - \eta_l^{j[2]}\right)^2 \middle| \mathcal{P}(t) \right], \tag{11}$$

is the segregation variance among the offspring of the parents $j[1]$ and $j[2]$ of individual j in generation t conditional only on the pedigree (not the traits), $\underline{\Sigma}_t^M$ is the minimum segregation variance of any family in the pedigree up to generation t, $\bar{\Delta}_t^M(\underline{z})$ is the maximum over the pedigree up to time $t-1$ of

$$\left| z^j - \frac{z^{j[1]} + z^{j[2]}}{2} \right|,$$

and

$$C\left(\sigma^2, |z|\right) = \frac{C'|z|}{\sqrt{\sigma_E^2 + \sigma^2}} + \frac{C''}{\sigma p(\sigma_E^2 + \sigma^2, |z|)} \left(1 + \frac{1}{\sigma^2}\right), \tag{12}$$

where $p(\sigma_E^2 + \sigma^2, x)$ is the density at x of a $\mathcal{N}(0, \sigma_E^2 + \sigma^2)$ distributed random variable. The constants C', C'' depend only on B, our bound on the scaled allelic effects.

In other words, we establish that the genetic components of the vector of traits in generation t follows the infinitesimal model with an error of order $t/\sqrt{M}$. Moreover, we see that the approximation breaks down if the segregation variance is too small in any family in our pedigree, or if the trait of an individual somewhere in the pedigree is too extreme in a way that is quantified by Eq. (12). Letting $M \to \infty$ in (11), $(\Sigma_t^M)_{jj}$ converges to $(\Sigma_t)_{jj}$, corresponding to $V_0(1 - F_j)$, where F_j is the probability of identity of the two parents of the j th individual, and so we recover Eq. (3).

We shall present an outline of the proof which is given in detail in the appendices. We start with the ancestors in the pedigree at generation zero and then work recursively through the pedigree. Crucially, we shall keep track of the rate of convergence to multivariate normality at each stage, as it is this that allows us to move from one generation to the next.

Mutation

To keep formulae simple, in the outline that follows we shall ignore mutation, but in the detailed derivation in the appendices we include house of cards mutation. Here we indicate how this changes the statement of the result. We suppose that the mutation probability per locus, per generation, is μ (independent of l). If there is a mutation at locus l, then the scaled allelic effect of the mutant is given by the random variable $\check{\eta}_l$ (also assumed to be bounded in modulus by B). We write $\bar{z}_\mu = \bar{z}_0 + \mathbb{E}[\sum_{l=1}^{M} \check{\eta}_l/\sqrt{M}]$ and

$$\check{\sigma}_M^2 = \frac{1}{M} \sum_{l=1}^{M} \mathrm{Var}(\check{\eta}_l),$$

and we suppose that $\check{\sigma}_M^2 \to \check{\sigma}^2$ as M tends to infinity.

The convergence in (10) is now of

$$\left(Z^j - \mu \bar{z}_\mu - (1-\mu)\frac{Z^{j[1]} + Z^{j[2]}}{2}\right)_{j=1,\dots,N_t}.$$

The segregation variance of the family of parents of individual j takes the form

$$(\Sigma_t^{M,\mu})_{jj} = (1-2\mu)(\Sigma_t^M)_{jj} + 2\mu \frac{1}{M} \sum_{l=1}^{M} \frac{\mathbb{E}[(\hat{\eta}_l - \check{\eta}_l)^2]}{2}, \tag{13}$$

where, since μ (the mutation rate per locus per generation) is typically very small we have dropped terms of order μ^2, and $\bar{\Delta}_t^M(\underline{z})$ is replaced by $\bar{\Delta}_t^{M,\mu}(\underline{z})$, the maximum over the pedigree up to time t of

$$\left| z_j - \mu \bar{z}_\mu - (1-\mu)\frac{z_j[1] + z_j[2]}{2} \right|.$$

3.1.1. Genetic component of the trait distribution in generation zero

The first step in our derivation is to show that as the number of loci tends to infinity, the distribution of the genetic component $(Z^1, \dots, Z^{N_0})$ of the traits in generation zero converges to that of a multivariate normal with mean vector $(\bar{z}_0, \dots, \bar{z}_0)$ and variance–covariance matrix $\hat{\sigma}^2 \mathrm{Id}$, where Id is the identity matrix.

To prove this, it is enough to show that for any choice of $\beta = (\beta_1, \dots, \beta_{N_0}) \in \mathbb{R}^{N_0}$,

$$\sum_{j=1}^{N_0} \beta_j Z^j \to Z_\beta,$$

where Z_β is normally distributed with mean $\bar{z}_0 \sum_{j=1}^{N_0} \beta_j$ and variance $\hat{\sigma}^2 \sum_{j=1}^{N_0} \beta_j^2$. Of course, the result will follow from the Central Limit Theorem, but we need to have some control over the *rate* of this convergence if we are to pass from one generation to the next. Moreover, although in the ancestral population the allelic effects

$\widehat{\eta}_l^i$ are independent, they do not all have the same distribution, and so we need an extension of the classical Central Limit Theorem. The version that we use is Theorem A.2, due to Rinott (1994). It is convenient to write $\|\beta\|_1 = \sum_{j=1}^{N_0} |\beta_j|$ and $\|\beta\|_2^2 = \sum_{j=1}^{N_0} \beta_j^2$. For generation zero, Theorem A.2 yields

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^{N_0} \beta_i (Z^i - \bar{z}_0)}{\|\beta\|_2 \widehat{\sigma}_M} \le z \right] - \Phi(z) \right|$$
$$\le \frac{C}{\|\beta\|_2 \sqrt{M} \widehat{\sigma}_M} \left(1 + \frac{1}{\|\beta\|_2^2 \, \widehat{\sigma}_M^2} \right) \qquad (14)$$

where Φ is the cumulative distribution function of a standard normal random variable and the constant C has an explicit expression (depending only on B, the bound that we imposed on the scaled allelic effects, and $\|\beta\|_1$). The details of the proof are in Appendix B. In particular, taking $\beta_k = 0$ for $k \neq j$ and $\beta_j = 1$, we read off the rate of convergence to the normal distribution of Z^j as the number of loci tends to infinity.

3.1.2. Strategy of the derivation

Our proof will be recursive. Suppose that we have our result for generation $(t-1)$. The key step is then to show that for individual j in generation t, conditioning on knowing $\mathcal{P}(t)$ and $\widetilde{Z}(t-1)$ provides negligible information on the values $\eta_l^{j[1]}$, $\eta_l^{j[2]}$ of the scaled allelic effects of locus l in its parents. Through an application of Bayes' rule, just as in our toy example at the beginning of the section, this will essentially boil down to showing that

$$\left| \frac{\mathbb{P}\left[\widetilde{Z}^{j[1]} = z_1 \big| \eta_l^{j[1]} = x, \mathcal{P}(t) \right]}{\mathbb{P}\left[\widetilde{Z}^{j[1]} = z_1 \big| \mathcal{P}(t) \right]} - 1 \right| \le \frac{t}{\sqrt{M}} C\left(\underline{\Sigma}_t^M, \bar{\Delta}_t^M(\underline{z}) \right), \quad (15)$$

where, since $\widetilde{Z}$ is a continuous random variable, the ratio on the left should be interpreted as a ratio of probability density functions, and the quantity C on the right was defined in (12). The proof depends crucially on knowing the rate of convergence of the distribution of the parental trait values to a multivariate normal.

What (15) allows us to deduce (via Bayes' rule) is that knowing the trait of an individual gives very little information about the allelic state at a single locus. Although intuitively clear, since all loci have a small effect on the trait, this is slightly delicate and, indeed, as we saw in our toy example at the beginning of the section (c.f. (12)), will break down if the segregation variance somewhere in our pedigree is small or if a trait in the pedigree is too extreme. Armed with (15), we can approximate the distribution of the allelic effects conditioned on $\mathcal{P}(t)$ and $\tilde{Z}(t-1)$ by those conditioned just on $\mathcal{P}(t)$ and then it is an easy matter to identify the limiting variance–covariance matrix of the random variables $(R^j)_{j=1,\ldots,N_t}$ (that we defined in (9)) in generation t.

Convergence of the vector of the genetic components of the trait values in generation t to a multivariate normal is then an application of Theorem A.2. Knowing the rate of this convergence allows us to prove the analogue of (15) for generation $(t+1)$, and so on.

3.1.3. One generation of reproduction

We have already proved the asymptotic normality for generation zero. To begin the recursion, we consider the first round of mating.

We suppose that, for each j, we know the parents of individual j and their observed trait values, $\widetilde{Z}^{j[1]}, \widetilde{Z}^{j[2]}$. In the notation defined above, this is precisely $\mathcal{P}(1)$ and $\widetilde{Z}(0) = \underline{z}$. Then we claim that *knowing this information*,

$$\left(Z^j - \frac{Z^{j[1]} + Z^{j[2]}}{2} \right)_{j=1,\ldots,N_1}$$

converges in distribution to a mean zero multivariate normal random variable with diagonal variance–covariance matrix Σ_1, whose on-diagonal entries are given by $(\Sigma_1)_{jj}$, the segregation variance among offspring of the parents of j th individual. By definition, for a given individual j in the first generation, we have

$$Z^j = \bar{z}_0 + \frac{1}{\sqrt{M}} \sum_{l=1}^{M} \left\{ X_l^j \eta_l^{j[1]} + (1 - X_l^j) \eta_l^{j[2]} \right\}$$

and so

$$R_j = \frac{1}{\sqrt{M}} \sum_{l=1}^{M} \left(\left(X_l^j - \frac{1}{2} \right) \eta_l^{j[1]} + \left((1 - X_l^j) - \frac{1}{2} \right) \eta_l^{j[2]} \right).$$

Since X_l^j is independent of $\mathcal{P}(1)$ and $\widetilde{Z}(0)$, the random variable R^j satisfies $\mathbb{E}[R^j | \mathcal{P}(1), \widetilde{Z}(0)] = 0$. We must calculate its variance. First, we use the normal approximation to the distribution of ancestral traits to show that

$$\left| \mathbb{P}\left[\eta_l^{j[1]} = x | \mathcal{P}(1), \widetilde{Z}(0) = \underline{z} \right] - \mathbb{P}[\eta_l^{j[1]} = x] \right|$$
$$\le \frac{1}{\sqrt{M}} C(\widehat{\sigma}_M^2, \bar{\Delta}_1^M(\underline{z})),$$

where C was defined in (12). Since individuals in the ancestral population are assumed to be unrelated, $\eta_l^{j[1]}$ and $\eta_l^{j[2]}$ are independent (provided the parents are distinct), and combining the calculation above with the symmetric one for $\eta_l^{j[2]}$ we can calculate that (for some $\alpha \in [-1, 1]$)

$$\left(\Sigma_1^M \right)_{jj} := \mathbb{E}\left[\left(R^j \right)^2 \big| \mathcal{P}(1), \widetilde{Z}(0) = \underline{z} \right]$$
$$= \mathbb{E}\left[\left(R^j \right)^2 \big| \mathcal{P}(1) \right] + \frac{\alpha}{\sqrt{M}} C\left(\widehat{\sigma}_M^2, \Delta\left(z_j \right) \right)$$
$$\to \frac{1}{4} \lim_{M \to \infty} \frac{1}{M} \sum_{l=1}^{M} \mathbb{E}\left[\left(\eta_l^{j[1]} - \eta_l^{j[2]} \right)^2 \right],$$

which is the limit of (11) with $t = 1$. The details are in Appendix D. Since the Bernoulli random variables that describe inheritance in different individuals are independent, it is easy to check that $\mathbb{E}[R^i R^j | \mathcal{P}(1), \widetilde{Z}(0) = \underline{z}] = 0$.

To verify convergence to a multivariate normal, we mimic what we did in the ancestral population: for an arbitrary vector $\beta = (\beta_1, \ldots, \beta_{N_1})$ we show that $\sum_{j=1}^{N_1} \beta_j R^j$ converges to a normal random variable as $M \to \infty$. The details (including the effects of mutation) are in Appendix D.

3.1.4. Generation t

We now proceed to the general case. We want to show that conditionally on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1) = \underline{z}$, R^j given by (9) converges in distribution as $M \to \infty$ to a mean zero, normally distributed random variable with diagonal variance–covariance matrix Σ_t given by the limit of (11). Independence of the Bernoulli random variables that determine inheritance in different individuals once again guarantees that for $i \neq j$, $\mathbb{E}[R^i R^j | \mathcal{P}(t), \widetilde{Z}(t-1) = \underline{z}] = 0$. As in generation one, the key is to show that

$$(\Sigma_t^M)_{jj} := \mathbb{E}\left[(R^j)^2 \big| \mathcal{P}(t), \widetilde{Z}(t-1) = \underline{z} \right]$$

is almost independent of $\underline{z}$. Convergence to $(\Sigma_t)_{jj}$ given by the limit of the expression in (11) is then straightforward (see Appendix E).

The involved step is to estimate

$$\mathbb{P}\left[\eta_l^{j[1]} = x | \mathcal{P}(t), \widetilde{Z}(t-1) = \underline{z} \right],$$

which, again by Bayes' rule, reduces to checking (15). At first sight, it seems that knowing the trait value of all the pedigree ancestors of the j th individual should give us much more information about η_l^j than just knowing the parental traits gave us in generation one.

In Appendix E we show that this is not really the case. The key is that the *differences* in trait values between individuals that are identical by descent at locus l are independent of the scaled allelic effect at locus l, so we do not accumulate any more information by observing all of these individuals than by observing just one of them.

We can perform entirely analogous calculations for the conditional joint law of $\eta_l^{j[1]}$ and $\eta_l^{j[2]}$. This enables us to identify the mean and the variance of the limiting distribution of traits in the population and, once again, Theorem A.2 can be used to establish that it is indeed a multivariate normal. In particular,

$$\left| \mathbb{P}\left[\frac{Z^j - \frac{Z^{j[1]}+Z^{j[2]}}{2}}{\sqrt{(\Sigma_t^M)_{jj}}} \le y \middle| \mathcal{P}(t), \widetilde{Z}(t-1) = \underline{z} \right] - \Phi(y) \right|$$
$$\le \frac{t}{\sqrt{M}} C\left(\underline{\Sigma}_t^M, \bar{\Delta}_t^M(\underline{z})\right).$$

In other words, we have established that the genetic components of the vector of traits follows the infinitesimal model with an error of order $1/\sqrt{M}$ per generation.

Adding an environmental component to the observed traits in the population makes the model more realistic; it also serves a mathematical purpose. As we explain in Remark C.1, without environmental noise, some extra conditions are required to guarantee a rate of convergence of order $1/\sqrt{M}$ (which is the best possible) to the limiting Gaussian distribution. If they are satisfied, then the calculations that we have just performed are unchanged if we condition on $Z(t-1)$, the genetic components of the trait values of individuals in the pedigree up to time $(t-1)$, instead of the observed values. Our assumption that the environmental noise is Gaussian is certainly unnecessarily restrictive (it would serve the same mathematical purpose if it had any smooth density). It has the advantage that we will be able to obtain explicit formulae for the distribution of observed traits.

3.1.5. Observed traits

In the presence of environmental noise, we cannot directly observe the genetic component of the trait. The infinitesimal model as stated above does not apply to the observed trait values. To see why, we write

$$\left(\Delta\widetilde{Z}^j\right)_{j=1,\dots,N_t} := \left(\widetilde{Z}^j - \mu\bar{z}_\mu - (1-\mu)\frac{\widetilde{Z}^{j[1]} + \widetilde{Z}^{j[2]}}{2}\right)_{j=1,\dots,N_t}$$
$$= \left(R^j + E^j - (1-\mu)\frac{E^{j[1]} + E^{j[2]}}{2}\right)_{j=1,\dots,N_t},$$

where $\mu = 0$ in the case with no mutation considered above. We have already checked that $(R^j)_{j=1,\dots,N_t}$ is a multivariate Gaussian vector which is (almost) independent of $\widetilde{Z}(t-1)$, and by assumption the same holds true for $(E^j)_{j=1,\dots,N_t}$. The difficulty is that the environmental components $E^{j[1]}$, $E^{j[2]}$ are *not* independent of the observed traits in generation $(t-1)$. However, under our assumption that the environmental noise is normally distributed, $(\Delta\widetilde{Z}^j)_{j=1,\dots,N_t}$ is still asymptotically normally distributed and we can derive recursions for the mean vector and variance–covariance matrices.

In what follows we assume that we are already in the asymptotic regime in which the genetic components of the traits follow a multivariate normal distribution. In fact we are accumulating errors of order $1/\sqrt{M}$ per generation in so doing.

To find the distribution of the observed traits in generation $(t+1)$ conditional on $\mathcal{P}(t+1)$ and $\widetilde{Z}(t)$, we need to calculate the conditional distribution of the vector $(E^j)_{j=1,\dots,N_t}$. Evidently this vector is independent of $\widetilde{Z}(t-1)$ and, granted that we have already calculated the corresponding conditional distributions for the environmental noise in generation $(t-1)$, calculating the conditional

distribution of $(E^j)_{j=1,\dots,N_t}$ is reduced to applying standard results on conditioning multivariate normal random variables on their sum. In particular, the conditioned vector is still a multivariate normal.

We write $E_c(t) = (E_c^j(t))_{j=1,\dots,N_t}$ for the random vector

$$(E^1, \dots, E^{N_t}) \middle| \mathcal{P}(t), \widetilde{Z}(t).$$

Notice that in contrast to what went before, we are conditioning on knowing all observed trait values up to and *including* generation t. We derive a recursion for the mean vectors $A_c(t)$ and the variance–covariance matrices $\Sigma_c^E(t)$ of these conditioned random vectors. Since environmental noise is not transmitted from parent to offspring, this is considerably more straightforward than our previous recursions.

In order to keep track of generations, we suppose that in generation t we condition on $\widetilde{Z}^j = \widetilde{z}^j(t)$. In this notation, the mean of $E_c^j(0)$ is determined by

$$A_c^j(0) = \left(\frac{\sigma_E^2}{\widehat{\sigma}^2 + \sigma_E^2}\right)(\widetilde{z}^j(0) - \bar{z}_0),$$

and the variance–covariance matrix of $E(0)$ is

$$\Sigma^E(0) = \left(\frac{\widehat{\sigma}^2 \sigma_E^2}{\widehat{\sigma}^2 + \sigma_E^2}\right) \text{Id}.$$

Now suppose that we have calculated $A_c(t-1)$ and $\Sigma_c^E(t-1)$. Then to find $A_c(t)$, $\Sigma_c^E(t)$, we first calculate the mean vector and the variance–covariance matrix for $\Delta\widetilde{Z}(t)$ conditional on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1)$. Denoting the corresponding quantities by $M(t)$ and $\Sigma_{BB}(t)$ respectively, the recursion which allows us to pass to the next generation reads

$$A_c(t) = \left(\sigma_E^2(\Sigma_{BB}(t))^{-1}\left(\Delta\widetilde{z}^j(t) - M(t)\right)\right)_{j=1,\dots,N_t},$$

and, for the variance–covariance matrix,

$$\Sigma_c^E(t) = \sigma_E^2\text{Id} - \sigma_E^4(\Sigma_{BB}(t))^{-1}.$$

The details of this derivation can be found in Appendix F. Although not as simple as the expressions one obtains for the genetic component of the trait alone, one can now read off the multivariate normal distribution of the observed trait values in generation t conditional on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1)$. Notice, in particular, that there will be correlations among individuals.

Accuracy of the infinitesimal model as an approximation

The infinitesimal model does not just say that the trait distributions in the population can be approximated by a multivariate normal random variable, but it also asserts that the variance of the genetic components of the traits is approximately independent of the trait values of the parents. What our calculations show is that in approximating Σ_t^M conditional on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1)$ by the same quantity conditioned only on $\mathcal{P}(t)$ (that is the right hand side of (11)), we are making an error of order (recall (15) in particular)

$$\frac{t}{\sqrt{M}} C\left(\underline{\Sigma}_t^M, \bar{\Delta}_t^M\right).$$

In other words, the infinitesimal model remains valid for $\mathcal{O}(\sqrt{M})$ generations, provided that the minimum segregation variance in the pedigree is not too small and none of the traits are too extreme, where both of these caveats are quantified by Eq. (12).

3.2. Beyond additivity: epistasis

In this section we outline how we can extend the infinitesimal model to capture some forms of epistasis. Unlike the additive setting, the mean value of the genetic component of the trait

across a family will no longer be at the mean value of the genetic components of the parents' traits as it will also depend on epistatic components. In this sense, the classical infinitesimal model cannot apply. However, under some conditions, it will still be the case that the variance of the trait distribution within a family will depend only on the pedigree (and the segregation variance in the ancestral population) and not on the observed traits.

We illustrate by considering only pairwise epistatic effects. We also ignore mutation and environmental noise. We write the trait in individual j as

$$Z^j = \bar{z}_0 + \frac{1}{\sqrt{M}} \sum_{l=1}^{M} \eta_l^j + \frac{1}{M} \sum_{1 \le l < m \le M} \eta_{lm}^j. \tag{16}$$

First observe that it cannot always be the case that the epistatic component of the trait is asymptotically normal. To see why, just take an additive trait, Z_A, which is asymptotically normally distributed and square it. Then $(Z_A)^2$ is of the form (16), but obviously does not have an asymptotically normal distribution.

On the other hand, one can write down easily interpretable conditions under which a quadratic form will be asymptotically normal. We do this before turning to general pairwise epistatic effects (for which writing down a precise result is more mathematically involved). Suppose that $X_1, X_2, \ldots, X_M$ are independent identically distributed random variables with mean zero, unit variance and finite fourth moment. Define

$$Z_{A(2)} = \sum_{1 \le l \le m \le M} a_{lm} X_l X_m,$$

for some real constants a_{lm}. We set $a_{lm} = a_{ml}$ and write A_M for the real symmetric matrix $A_M = (a_{lm})_{1 \le l, m \le M}$. We assume that

$$\lim_{M \to \infty} \sigma_M^{-4} \mathrm{Tr}(A_M^4) = 0, \quad \lim_{M \to \infty} \sigma_M^{-2} \max_l \sum_{m=1}^{M} a_{lm}^2 = 0, \tag{17}$$

where $\sigma_M^2 = \frac{1}{2} \mathrm{Tr}(A_M^2)$ is the variance of $Z_{A(2)}$. The first condition ensures that the *fourth moment condition* is satisfied, that is

$$\mathbb{E}\left[\frac{Z_{A(2)}^4}{\sigma_M^4} \right] \to 3 \qquad \text{as } M \to \infty.$$

The second condition ensures that no single locus has too much influence on the value of the trait. Under these conditions, Proposition 3.1 of Chatterjee (2008) tells us that the Kantorovich–Wasserstein distance (see Appendix A) between the distribution of $Z_{A(2)}/\sigma_M$ and that of a standard normal distribution is bounded above by

$$\left(\frac{\mathrm{Tr}(A_M^4)}{2\sigma_M^4} \right)^{\frac{1}{2}} + \frac{5}{2\sigma_M^3} \sum_{l=1}^{M} \left(\sum_{m=1}^{M} a_{lm}^2 \right)^{3/2}.$$

If σ_M^2 is order one and we assume further that $|a_{lm}| \le B/M$ uniformly in l, m, and $|\sum_{m=1}^{M} a_{lm}| \le B/\sqrt{M}$ uniformly in l, for some constant B, which ensures that the total allelic effect of locus l is bounded in the same way as in our analysis of the additive case, then the rate of convergence is of order $1/\sqrt{M}$.

The form $a_{lm} X_l X_m$ is not particularly natural for the epistatic effect of the loci l and m, but this result can be extended. For M unlinked loci, the genetic component of an arbitrary trait can be expressed as a function $f(\chi_1, \ldots, \chi_M)$ where the χ_l are independent (not necessarily identically distributed) random variables representing the underlying allelic states. In Appendix A, we explain the Hoeffding decomposition of a general function of this form, but here we focus on functions corresponding to traits with additive and pairwise epistatic effects only.

First recall that for a function $f_{lm}(\chi_l, \chi_m)$, the conditional expectation $\mathbb{E}[f_{lm}(\chi_l, \chi_m)|\chi_l]$ is a function of χ_l, and similarly $\mathbb{E}[f_{lm}(\chi_l, \chi_m)|\chi_m]$ is a function of χ_m. Writing

$$\alpha_{lm}(\chi_l, \chi_m) = f_{lm}(\chi_l, \chi_m) - \mathbb{E}[f_{lm}(\chi_l, \chi_m)|\chi_l]$$
$$- \mathbb{E}[f_{lm}(\chi_l, \chi_m)|\chi_m] + \mathbb{E}[f_{lm}(\chi_l, \chi_m)],$$

any trait involving pairwise epistatic interactions can be written in the form

$$Z = \bar{z}_0 + \sum_{l=1}^{M} \alpha_l + \sum_{1 \le l \le m \le M} \alpha_{lm},$$

where

$$\mathbb{E}[\alpha_{lm}|\chi_l] = 0 = \mathbb{E}[\alpha_{lm}|\chi_m]. \tag{18}$$

The sum over α_{lm} is called a degenerate U-statistic of order two and a Central Limit Theorem due to de Jong (1990) provides conditions under which it is asymptotically normally distributed as $M \to \infty$. Döbler and Peccati (2017) establish the rate of convergence in the de Jong Central Limit Theorem. Some more detail is given in Appendix A. To apply their result in this context, let us write $\sigma_{lm}^2 = \mathrm{Var}(\alpha_{lm})$ and $\sigma_M^2 = \mathrm{Var}(\sum_{1 \le l \le m \le M} \alpha_{lm})$, then provided

$$\rho(M)^2 := \sigma_M^{-2} \max_{1 \le l \le M} \sum_{m=1}^{M} \sigma_{lm}^2 \to 0 \qquad \text{as } M \to \infty,$$

and

$$\mathbb{E}\left[\left(\frac{1}{\sigma_M} \sum_{1 \le l \le m \le M} \alpha_{lm} \right)^4 \right] \to 3 \qquad \text{as } M \to \infty,$$

de Jong' result tells us that the sum $\sum_{1 \le l \le m \le M} \alpha_{lm}$ is asymptotically normal. Just as in our example above, if we suppose that σ_M^2 is order 1 and $\alpha_{lm} = \eta_{lm}/M$ with $|\eta_{lm}|$ and $|\sum_{m=1}^{M} \eta_{lm}/\sqrt{M}|$ uniformly bounded, then one can check that the rate of convergence is order $1/\sqrt{M}$. In fact the results of Döbler and Peccati (2017) are much more general than this and would allow us to state an analogous result for a trait involving higher order epistatic interactions, but the statement of the result becomes more mathematically involved.

The justification of the infinitesimal model follows a familiar pattern. What we have so far is enough to prove asymptotic normality in generation 0. The analogue of (15) follows essentially exactly as before. By analogy with what went before we write

$$\Delta Z^j = \frac{1}{\sqrt{M}} \sum_{l=1}^{M} \left(\eta_l^j - \frac{\eta_l^{j[1]} + \eta_l^{j[2]}}{2} \right)$$
$$+ \frac{1}{M} \sum_{1 \le l \le m \le M} \left(\eta_{lm} - \frac{\eta_{lm}^{j[1]} + \eta_{lm}^{j[2]}}{4} \right).$$

But when we pass to later generations, we encounter two complications. First, it is no longer the case that $\mathbb{E}[\Delta Z^j] = 0$. Nonetheless, as we explain in more detail in Appendix G, a coupling argument, which exploits the fact that conditional on the parental traits, the allelic states have almost the same distribution as the unconditioned one, combined with (18) implies that its expectation is order $1/\sqrt{M}$.

The second complication is that the ΔZ^j's are no longer asymptotically uncorrelated. In the first generation, correlation comes about if two individuals share the same parents, are identical by descent at two distinct loci $\{l, m\}$ *and* neither individual inherited the alleles at both l and m from the same parent. To see this, the chance that they did inherit both loci from the same parent, $j[1]$ say, is $1/4$ and then this term is cancelled by the corresponding centring term, $\eta_{lm}^{j[1]}$. If for both individuals the two loci are inherited

from different parents, then no such cancellation takes place. In particular, such correlation only arises because there are two loci involved in η_{lm}, which is why it did not appear in the additive case. In later generations, correlation again comes from identity at two distinct loci, but there may be multiple different routes through the pedigree that result in this, so we no longer require that $i[1]$ and $i[2]$ are the same as $j[1]$ and $j[2]$. However, just as in generation one, for a given pair of loci $\{l, m\}$, the contribution to the correlation will be zero if the individuals share no pedigree ancestry or if one of the individuals inherited both loci from the same parent. Thus the contribution from the loci $\{l, m\}$ takes the form

$$\frac{1}{4}\text{Var}(\widehat{\eta}_{lm})F_2(i[1], i[2], j[1], j[2])$$

$$:= \frac{1}{2}\text{Var}(\widehat{\eta}_{lm})\{F(i[1], j[1])F(i[2], j[2])$$
$$+ F(i[2], j[1])F(i[1], j[2])\},$$

where $F(a, b)$ is the probability that individuals a and b are identical by descent at a given locus and again we have exploited the fact that, knowing the pedigree, the pattern of inheritance is almost unchanged by knowing trait values. Summing over pairs of loci $\{l, m\}$, we obtain that for $i \neq j$, (in the limit as $M \to \infty$),

$$(\Sigma_t)_{ij} = \frac{1}{2}F_2(i[1], i[2], j[1], j[2])V_{A(2)}(0).$$

A similar calculation shows that, in the limit as $M \to \infty$, the epistatic component contributes

$$\left(1 + \frac{1}{8}(1 + F(j[1], j[2])^2) - \frac{1}{4}(1 + F(j[1], j[2]))^2\right)V_{A(2)}(0)$$

to $(\Sigma_t)_{jj}$. Although notationally complicated, the ideas are the same as in the additive case and we can conclude that conditionally on $Z(t-1)$ and $\mathcal{P}(t)$, the vector $(\Delta Z^1(t), \dots \Delta Z^{N_t}(t))$ converges in distribution as $M \to \infty$ to a multivariate Gaussian with mean zero and variance–covariance matrix Σ_t. Armed with the speed of convergence (which is again order $1/\sqrt{M}$) and Bayes' rule, we can prove that conditioning on trait values (in fact even conditioning on z_A, $z_{A(2)}$) provides little information about the allelic state at any particular locus and then use the unconditioned distribution of allelic states to estimate the variance–covariance matrix for the next generation, and so on.

3.3. Numerical examples

In this section, we illustrate our results with some numerical examples. We focus on cases with no environmental noise or mutation. We begin with an example that shows how the genetic variance is much less sensitive to selection than the mean. We then show how the genetic variance scales with the number of loci, and finally, show that the variance amongst offspring depends only weakly on the parents' trait values.

All our examples are based on simulations of a Wright–Fisher model of 100 haploid individuals. With the exception of Fig. 4, which investigates how rapidly the infinitesimal limit is approached as the number of loci increases, we set the number of loci $M = 1000$. Our choice of parameters is constrained by computational limits. We mostly simulated equal 'main' effects, corresponding to $|\eta_l|$ in our derivation of Section 3 all taking the same constant value, because the infinitesimal limit is then approached with a minimal number of loci. Exponentially distributed effects might be more realistic, but require an order of magnitude more loci to approach the infinitesimal limit (see Fig. 4).

Simulation of larger populations would be challenging, if very large numbers of loci are followed. (As noted above, simulation of the infinitesimal is possible for large N provided that the relatedness is approximated as the same for all pairs.) Following

Robertson (1960), we expect that the infinitesimal model will break down when Ns per locus becomes large.

We also consider pairwise epistasis. One situation in which we expect the infinitesimal model to be an accurate approximation is if the epistatic effects are sufficiently sparse. Our example is constructed in two stages. First, independently for each ordered pair (l, m) of loci with $l \neq m$, with probability $1/M$ we declare it to have a non-zero epistatic effect. Then, each non-zero interaction is assigned an independent value γ, sampled from a normal distribution with mean zero and standard deviation $4/\sqrt{M}$. With this construction $\gamma_{lm} \neq \gamma_{ml}$. The trait is now defined as $z = \delta \cdot \alpha + \delta \cdot \gamma \cdot \delta^T$, where the entries in the vector δ, which records the genotype, are $\pm 1/2$, the vector α records the absolute magnitude of the allelic effects at each locus and '·' is dot product. The epistatic and main effects were scaled with respect to the number of loci, M, so that both the additive and epistatic variances are of order 1.

We chose $N = 100$ haploid individuals, in order to be able to follow the matrix of identities. Since we have no mutation, variation then dissipates over about $N = 100$ generations. We ran simulations either with no selection, or with directional selection $\beta = 0.2$, so that fitness W is proportional to $e^{\beta z}$. Simulations were started with allele frequencies drawn from a Beta distribution with mean $p = 0.2$ and variance $0.2pq$.

With selection on a heritable trait, fitness is also heritable, which speeds up the loss of genetic variance due to random drift (Robertson, 1961); the variance declines in proportion to $1/N_e$ per generation, where N_e is somewhat less than the census number. However, in our examples, the variance in fitness, $\beta^2 V_G$, is small, and so the relatedness, F, is not appreciably increased by selection.

Fig. 2 shows how selection affects the mean and the components of the variance of a trait that is determined by $M = 1000$ loci, in a population of $N = 100$ haploid individuals. It is natural to work with the dimensionless quantities $z/\sqrt{V_A}$ and $\beta\sqrt{V_A}$ (where V_A is the initial additive variance). Selection on an additive trait changes the mean by 4.5 genetic standard deviations over 100 generations whilst the additive genetic variance decreases by a factor 0.355 (top row). This is almost the same as the decrease with no selection, 0.364, and close to the neutral expectation $(1 - 1/N)^{100} = 0.366$. In fact, selection is expected to alter the pedigree and thus the rate of loss of diversity, but the effect actually seems very small. The bottom row shows an example with sparse pairwise epistasis. The additive variance is much higher, and the mean now changes by $10.9\sqrt{V_A}$ (where V_A is again taken at time 0) over 100 generations. In both the additive and epistatic examples, averaging the per-generation scaled s.d. $\beta\sqrt{V_A}$ over 100 generations gives a good prediction of the change in (scaled) mean per generation. With epistasis, but no selection, the additive variance decreases only by a factor 0.548 over 100 generations (upper dashed line at bottom right), because non-additive variance is 'converted' into additive variance. Now, selection does substantially reduce the additive variance, after about 30 generations.

A cornerstone of our derivation of the infinitesimal model was the result that the distribution of allele frequencies is hardly affected by selection. Fig. 3 shows the change in allele frequency over time. In the presence of epistasis, the marginal effect of an allele on the trait depends on current allele frequencies at other loci. In particular, alleles that have a positive effect on the trait in the initial population may have a different effect in the final generation. In order to determine whether the allele frequency spectrum is substantially biased towards alleles that have a positive overall effect on the trait, in the presence of epistasis we recalculate the marginal effects of each allele at the final generation and relabel them accordingly. Over 100 generations, random drift fixes alleles that are rare or common, leaving a flat distribution of frequencies of segregating alleles (grey lines at middle and bottom). Selection shifts the distribution in favour of alleles with a positive effect
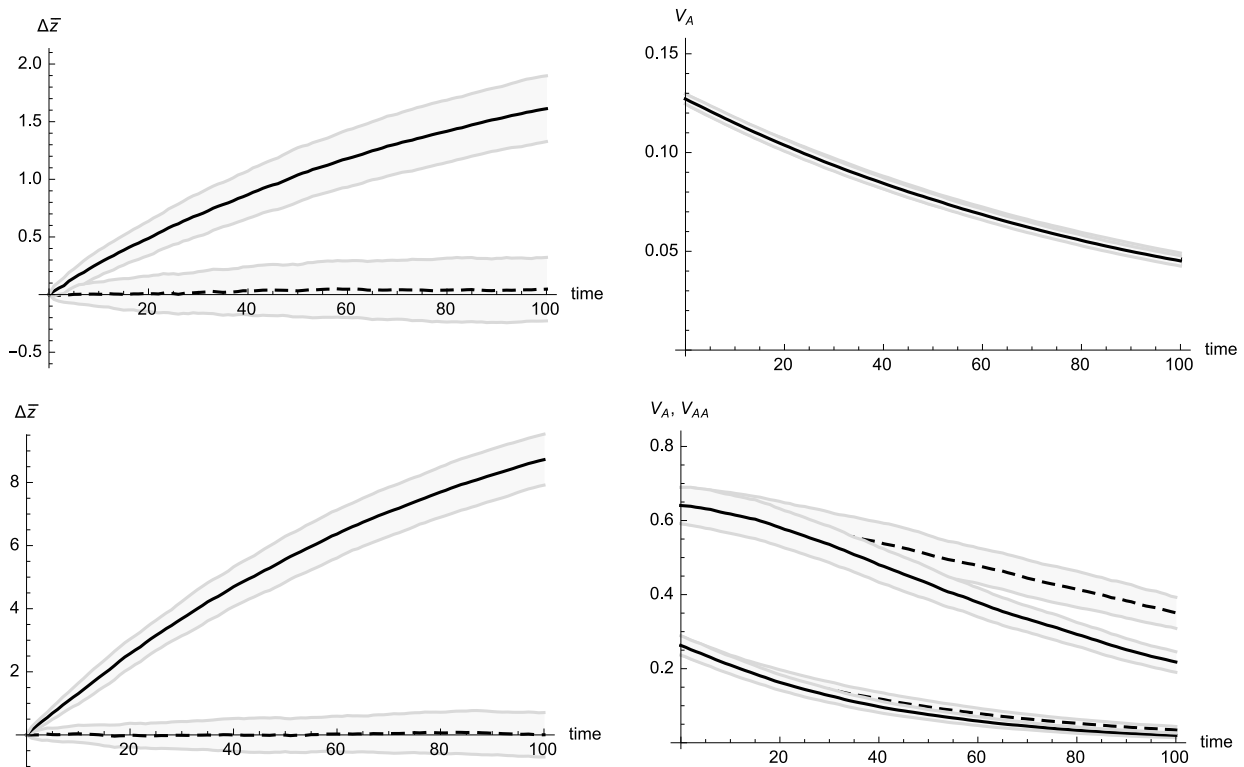
Fig. 2. The effect of selection on the mean and variance components. Each panel contrasts directional selection, $\beta = 0.2$, on a trait (solid line) with the neutral case (dashed line); shaded areas indicate ± 1 standard deviation. The top row shows the additive case, with equal effects but random sign, $\eta = \pm\frac{1}{\sqrt{M}}$; in this example, $M = 1000$ loci, $N = 100$ haploid individuals. The lower row shows an example of sparse pairwise epistasis, as described in the text. The left column shows the change in mean from its initial value, and the right column shows the additive variance, V_A, and with epistasis, the additive $\times$ additive variance, V_{AA} (lower pair of curves at bottom right). Only the genic components of variance are shown; random linkage disequilibria produce substantial fluctuations, but make no difference on average (results not shown). Initial allele frequencies are drawn from a U-shaped Beta distribution with mean $p = 0.2$ and variance $0.2pq$. Individuals are produced by Wright–Fisher sampling, from parents chosen with probability proportional to $W = e^{\beta z}$. For each example, ten sets of allelic and epistatic effects are drawn and, for each of those, ten populations are evolved; this gives 100 replicates in all.

on the trait. In the additive case, the shift is symmetrical, so that there is virtually no change in the additive genetic variance: $\mathbb{E}[pq]$ stays the same. In contrast, selection on an epistatic trait reduces the overall frequency of segregating alleles. This may be because with epistasis, allelic effects vary with allele frequency, so that alleles experience an additional random fluctuation that reduces diversity.

With no selection, the rate of decrease of variance, $-\partial_t \log(V_A)$, is close to $1/N$, independently of the number of loci (compare small dots with dashed line in Fig. 4). With selection on a small number of loci ($M = 30$; left of Fig. 4), the additive variance declines much faster, as favourable alleles are swept to fixation. The excess rate of decrease is inversely proportional to the number of loci: $-N\partial_t \log(V_A) - 1 = \mathcal{O}(1/M)$; the exponent on the relation, estimated by regression on a log–log scale, is -1.007, -1.019, -0.965 for the three examples of equal effects, exponentially distributed effects, and equal effects plus epistasis. Note that as in Fig. 2, the additive variance is much more sensitive to selection in the presence of epistasis (compare upper large dots with medium black dots). However, in both cases the additive variance scales as close to $1/M$.

This is a much faster approach to the infinitesimal model than the upper bound of $1/\sqrt{M}$ set by our mathematical results. By considering the argument at the beginning of Section 3 for the additive model with equal main effects, we can see that the rate $1/\sqrt{M}$ cannot be improved upon in our general result (which applies even when we consider individual families and condition on ancestral traits). However, when, as in Fig. 4, we consider the variance across

the whole population, we can expect faster convergence. We can understand why this is as follows. The genic component of the additive variance is $V_A = \sum_{l=1}^{M}\alpha_l^2 p_l q_l$. Since the rate of change of the allele frequency at locus l due to directional selection β on the trait is $\beta\alpha_l p_l q_l$, we have $-\partial_t V_A = \beta\sum_{l=1}^{M}\alpha_l^3 p_l q_l (p_l - q_l)$ (ignoring the change in the marginal allelic effect due to epistasis). We assumed that α_l has random sign, so that the expected initial rate of change is zero; indeed, when we sample replicates with different effects, α, the rate of decrease measured over the first 10 generations is closely correlated with $\beta\sum_{l=1}^{M}\alpha_l^3 p_l q_l (p_l - q_l)$, and fluctuates in sign (results for $M = 30$ not shown); the standard deviation of this initial rate scales as $1/M$. However, the rates of decrease measured over 100 generations are almost all positive, and vary much less than does the initial prediction $\beta\sum_{l=1}^{M}\alpha_l^3 p_{0,l} q_{0,l}(p_{0,l} - q_{0,l})$, based on the initial allele frequencies. This consistently positive rate of decrease arises because the allele frequencies p_l become correlated with the allelic effect, α_l: with no mutation, favourable alleles eventually become common, so that the additive variance decreases as they move to fixation.

For the rest of our examples, we simulated a population of $N = 100$ haploid individuals with $M = 1000$ loci. The main allelic effects at each locus are taken to be $1/\sqrt{M}$, and examples with sparse epistatic interactions are constructed exactly as above. Simulations were run for 100 generations, tracking the full pedigree. We then calculated the matrix of pairwise identity by descent, F, relative to the initial population, and also relative to generation 80. We assume that the probability that two pairs of genes at different loci are both identical by descent is F^2.
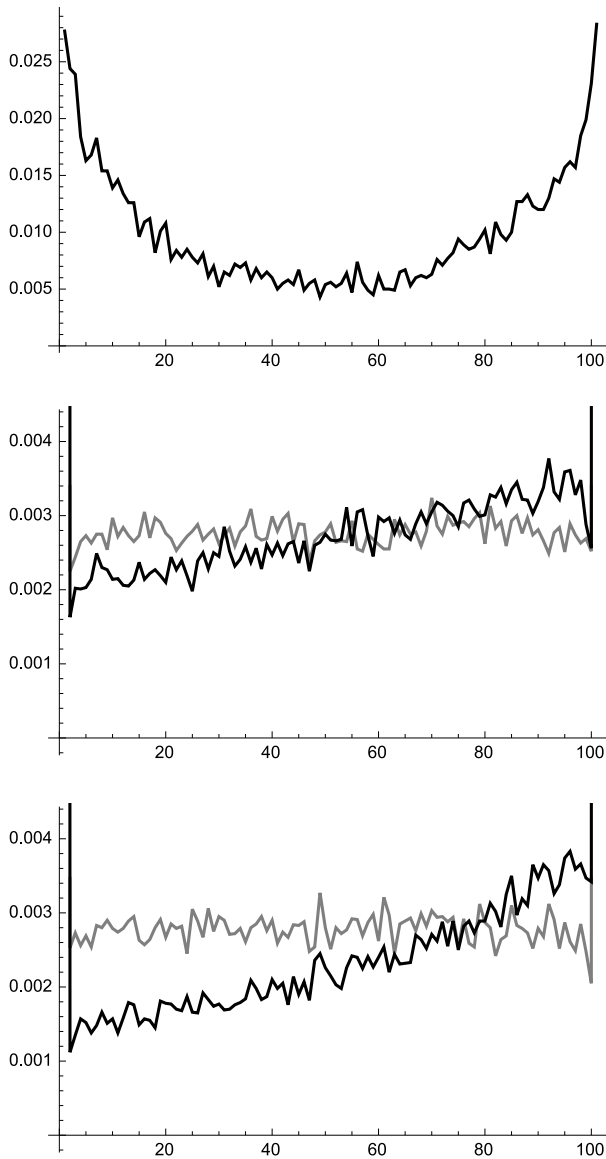
Fig. 3. Allele frequency distributions for the examples in Fig. 2. The horizontal axis shows the number of copies of the +allele $(0, \ldots, N)$ in a population of size $N = 100$. The vertical axis is the proportion of the $M = 1000$ loci with that number of + alleles. Top: Initial allele frequency distribution, which is independent of allelic effect. Middle: Additive case, after 100 generations, with no selection (grey) or with selection $\beta = 0.2$ (black). Bottom: With epistasis, after 100 generations where an allele is defined to be + if its marginal effect in the final generation is positive (see main text).

In Fig. 5, we plot the mean and variance within families of the additive and non-additive components against the average of the corresponding components in the two parents; parameters are as in the epistatic example shown in Fig. 2. We choose pairs of unrelated parents so that under the infinitesimal model the variances should be constant across pairs. Even in the presence of epistasis, the mean of the additive component in the offspring must be precisely the mean of the additive component in the parents (top left). The mean additive × additive component in the offspring equals half of the mid-parental value of that component, plus a random component that varies between pairs of parents. The slope of the regression line in the top right panel is 0.58, reasonably close to the expectation. The variance of the additive and non-additive components amongst offspring is independent of the mean parental components (bottom row).
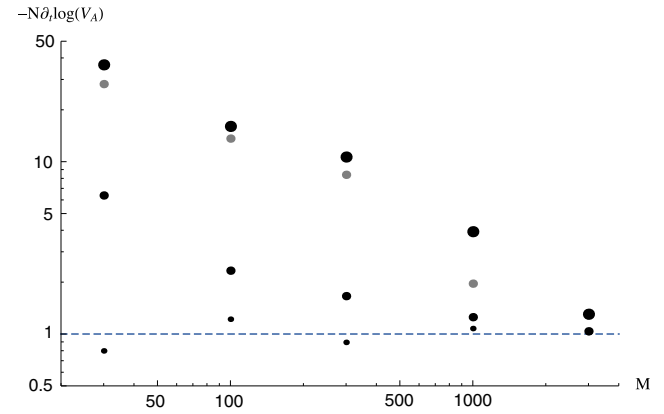

Fig. 4. The effect of selection on the genetic variance decreases with the number of loci, M. The rate of decrease of additive genetic variance, multiplied by population size, $N\partial_t \log(V_A)$, is plotted against M on a log–log scale. The neutral rate of decrease of variance is indicated by the dashed line at $N\partial_t \log(V_A) = 1$. Small dots show the neutral case, with equal and additive effects and random sign, $\alpha = \pm 1/\sqrt{M}$. Medium black dots show directional selection $\beta = 0.2$, again with equal and additive effects, $\alpha = \pm 1/\sqrt{M}$. Medium grey dots show $\beta = 0.2$, with additive effects drawn from an exponential distribution with mean $\alpha = 1/\sqrt{M}$ and random sign. Large dots show $\beta = 0.2$, with equal allelic effects $\alpha = \pm 1/\sqrt{M}$, plus sparse pairwise epistasis, as described in the text. Initial allele frequencies are drawn as in Fig. 2. The rate of decrease of variance is estimated by regression over 100 generations, V_A being averaged over 10 replicates. These replicates are evolved independently, starting from the same randomly chosen allelic effects, α, γ, and initial allele frequencies. Only the genic component of the additive variance is shown.

Finally, in Fig. 6, we plot the decline of the additive and non-additive variance with relatedness, F. Under the infinitesimal model, the within-family additive variance is $\frac{V_A}{2}(1 - F)$, where F is the probability of identity between the genomes that unite at meiosis. With sparse pairwise epistasis, the within-family non-additive variance is $\frac{V_{AA}}{4}(1 - F)(3 + F)$ (this expression can be obtained by the same analysis as in Section 3.2, where we compute the non-additive variance due to groups of genes inherited from different parents). Fig. 6 is based on a population of 100, simulated for 20 generations under selection $\beta = 0.2$, as for Fig. 5; the average relatedness is then $F = 0.18 \approx 1 - (1 - 1/N)^{20}$. The theoretical predictions shown by the solid curves are based on the additive and non-additive variance components in the base population. There is a small deviation from these predictions, because the observed genetic variances include random linkage disequilibria amongst the 100 sampled individuals, whereas the predictions are based on the allele frequencies in the base population.

4. Discussion

Typically, the distribution of a quantitative trait amongst offspring is Gaussian, with a mean intermediate between the parents' trait values, and a variance that is independent of those values. This observation goes back to Galton (1877), and was explained by Fisher (1918) as being the result of a large number of unlinked loci with additive effects. The variance amongst offspring depends on the relatedness of the parents, which can be predicted from the pedigree. This infinitesimal model thus provides a complete description of the short-term evolution of quantitative traits, which is independent of any knowledge of the genetics.

We have set out a rigorous justification for the infinitesimal model, which clarifies some of the conditions under which it holds. These are surprisingly general. In the additive case, we can include arbitrary selection and population structure, provided that the segregation variance is not too small and traits are not too extreme in a sense that is made precise by Eq. (12). The derivation
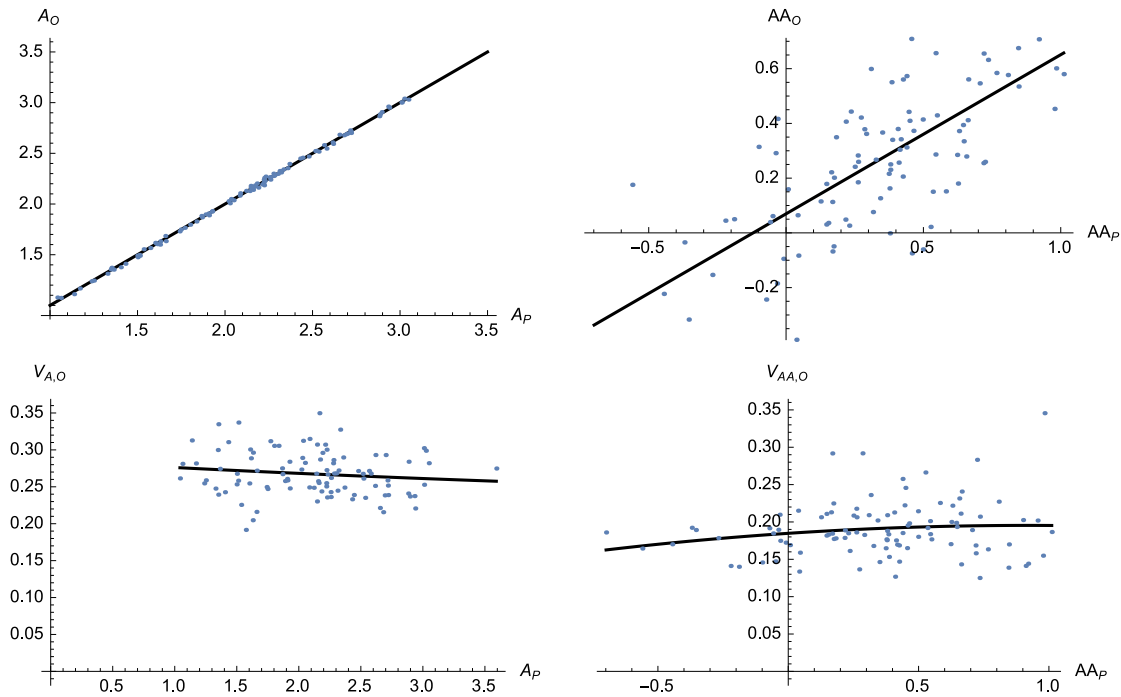
Fig. 5. The mean and variance of offspring traits, plotted against components of the parents' trait values. Top left: Additive component of offspring, A_O, against the mean of the parents' additive component, A_P. The line represents $A_O = A_P$. Top right: The same, but for the additive $\times$ additive components. The line shows a linear regression. Bottom left: Additive variance amongst offspring, $V_{A,O}$ against the mean additive components of the parents, A_P. Bottom right: Additive $\times$ additive variance of offspring against the mean additive $\times$ additive component of the parents. Lines in the bottom row show quadratic regressions. The example shows a non-additive trait under selection $\beta = 0.2$, with $M = 1000$ loci and $N = 100$ haploid individuals, as in Fig. 2 (bottom row). At generation 20, 100 pairs of minimally related parents ($F = 0.16$) were chosen, and 1000 offspring were generated for each pair. For each offspring, the components of trait value were calculated relative to the allele frequencies, p in the base population. Defining genotype by $X \in \{0, 1\}$, these components are $A = \zeta.(\alpha + (\gamma + \gamma^T).(p - \frac{1}{2}))$, $AA = \zeta.\gamma.\zeta^T$, where $\zeta = X - p$, and the allelic effects α, γ are drawn as for Fig. 2.
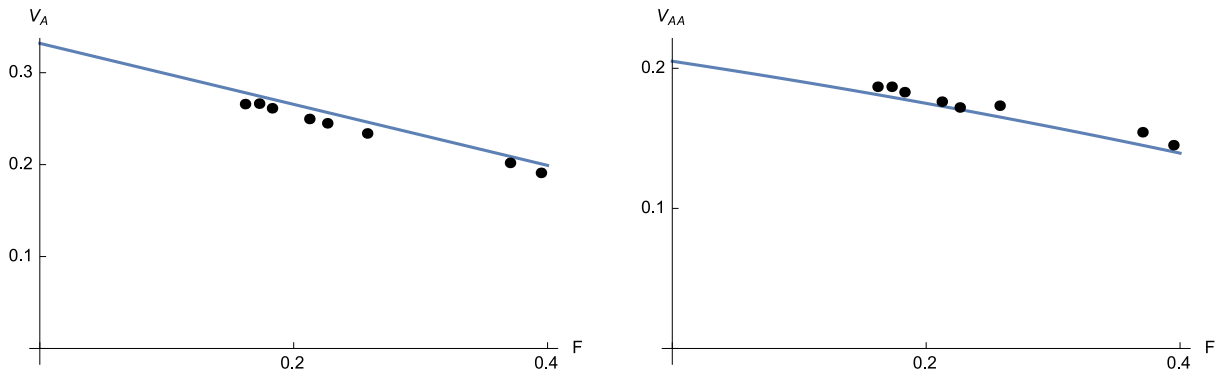
Fig. 6. Additive and non-additive variance amongst offspring declines with pairwise relatedness between parents. Solid lines show the theoretical expectations: $\frac{V_A}{2}(1 - F)$, $\frac{V_{AA}}{4}(1 - F)(3 + F)$. Values are based on a population of 100 individuals, generated as in Fig. 5. The 4950 distinct pairs of parents are pooled according to their relatedness, and the variance amongst offspring is estimated from 1000 offspring produced by each pair.

includes (house of cards) mutation and environmental noise. Most surprising, the argument that the variance of the distribution of the trait amongst offspring is insensitive to selection carries over to allow some forms of epistasis. With epistasis, we must now specify a set of variance components, which predict the variance amongst offspring on an arbitrary pedigree. In all cases, the mathematical analysis shows that the infinitesimal model holds up to an error which is at most of the order of $1/\sqrt{M}$, where M is the number of loci, while Fig. 4 suggests that in some cases the error could be as small as order $1/M$.

We have not considered dominance here. With dominance, the variance components no longer suffice to predict the offspring distribution: more complex quantities are involved (Barton and Turelli, 2004). Nonetheless, the proof of convergence of trait values on a given pedigree to a multivariate normal given by Lange (1978)

does include dominance and we anticipate that our central result still holds in the limit of a large number of loci: the variance of the traits among offspring is independent of the parents' trait values, and hence insensitive to selection. This, along with a more thorough mathematical investigation of the most general conditions under which epistatic interactions do not disrupt the infinitesimal model, will be the subject of future work.

We have assumed throughout that all loci are unlinked. This is of course inconsistent with assuming a very large number of loci on a finite genetic map. Linkage will reduce the rate at which segregation variance will be released, and more seriously, the breeding value and within-family variance are no longer sufficient to predict the evolution of the population: associations will build up between linked loci, and will only slowly be dissipated. However, in the limit of a long genetic map, selection still has negligible effect on the genetic variance (Bulmer, 1974; Santiago, 1998).

One can imagine an extension of the infinitesimal model to linear genomes, which can readily be implemented in an individual-based simulation. Imagine two long genomes of map length R, which initially have a certain total effect on the trait, $z_1([0, R])$, $z_2([0, R])$. Recombination between these genomes generates a gamete that is a mosaic of blocks derived from one or other parent, $\{z_1([0, x_1]), z_2([x_1, x_2]), \ldots\}$. Conditional on the breakpoints, the values associated with the segments $z_i([x_i, x_{i+1}])$, form a multivariate Gaussian, conditioned on the sum $z_i([0, R])$. At the level of the population, this model is implicit in studies of the effects of background selection, in which heritable fitness variance due to deleterious mutations is spread over the genome (e.g. Good and Desai, 2014).

The infinitesimal model requires that a sufficient number of loci contribute to the trait. With strong inbreeding, the number of contributing loci may become too small for the model to be accurate. This may be a particular problem if variance is contributed by rare recessive alleles: only a few such alleles may contribute in a cross between two close relatives. Thus, the infinitesimal model may break down under strong inbreeding between particular individuals or in particular subpopulations.

For how long can we expect the infinitesimal model to be accurate? The wide use of the model in animal breeding suggests that it is accurate (or at least, useful) for many tens of generations. Indeed, the sustained response to artificial selection that is typically seen is the strongest support for the infinitesimal approximation (Barton and Keightley, 2002). Remarkably, Weber and Diggins (1990, Fig. 4) found that for a wide range of traits and model organisms, the response to selection over 50 generations is close to the infinitesimal prediction. Responses tend to be slightly below the prediction, suggesting that selection is reducing the variance faster than expected by random drift, but the closeness of the fit implies that most of the selection response is due to alleles that are influenced mainly by random drift (i.e., that $N_e s < 1$ or less).

This evidence comes from relatively small populations, and short timescales. In the longer term, mutation becomes significant, and the infinitesimal model predicts a genetic variance in a balance between mutation and drift of $N_e V_m$ for a haploid population. This cannot plausibly explain observed heritabilities in large natural populations, since genetic variances do not show a strong increase with population size. (Though, we note that sequence diversity also shows a weaker increase with census population size than expected from naive neutral theory. It is not clear whether quantitative genetic variance increases in proportion with sequence diversity; Frankham, 1996; Willi et al., 2006.) It is widely believed that genetic variance is due to a balance between mutation and selection against deleterious mutations. However, it is not clear whether selection acts on the trait or on the pleiotropic effects of the alleles involved, and the contribution of balancing selection of various kinds is unknown (Johnson and Barton, 2005). The infinitesimal model may remain accurate at least for times shorter than $1/s$; however, the effects of selection at the underlying loci need further theoretical investigation. Estimates of the distribution of fitness effects (largely based on evidence from Drosophila) suggest that there may be a significant fraction of very weakly selected alleles (e.g. Loewe and Charlesworth, 2006); if these contribute to traits as well as to fitness, then the infinitesimal model may hold for long times. However, Charlesworth (2015) concluded that the quantitative genetics of fitness variation in Drosophila can only be reconciled with estimates of fitness effects from population genomics if most fitness variance is either due to relatively strongly selected mutations ($N_e s \gg 1$), or to the side-effects of balancing selection.

The enormous efforts put into mapping quantitative trait loci (QTL), and more recently, to finding associations between genome-wide markers and quantitative traits (GWAS), have identified many QTL, but typically, have not explained much of the genetic variance. There is no mystery about this "missing heritability": it is to be expected if genetic variance is due to large numbers of alleles of small effect. In addition, SNP markers may not be in complete association with causal alleles, especially if the latter are at more extreme frequencies (Yang et al., 2010). Thus, it may only be possible to identify the small fraction of individual alleles in the upper tail of the distribution of effects, even if the whole genome and the whole population are sequenced (Boyle et al., 2017). Nevertheless, a regression of trait on sequence can significantly improve predictions of breeding value, even when individual loci cannot be identified: this is the basis of "genomic selection" (Meuwissen et al., 2013). It may be that natural selection is in just the same position as a breeder: selection may change the mean rapidly and predictably, even when the probability distribution of any particular allele frequency is hardly perturbed.

Acknowledgements

Appendix A. Generalised central limit theorems

There is now a huge literature on rates of convergence in the Central Limit Theorem, mostly dependent upon Stein's method. Here we present two forms. We use the first, due to Rinott (1994), in the additive case. Although we do not use this property here, it allows some dependency between elements in the sum, which would be useful if we wanted to think of loci grouped on chromosomes, for example. From our perspective it is convenient as, not only does it allow non-identically distributed summands, but also we can apply it without directly checking a fourth moment condition. The second result, due to Döbler and Peccati (2016) develops quantitative central limit theorems for degenerate U-statistics that do not have the form of homogeneous sums. This is what is required to go beyond the somewhat contrived example in the main text in which pairwise epistatic effects are determined by a quadratic form and are therefore amenable to the analysis (developed in a completely different setting) of Chatterjee (2008). Although this result would apply in the additive case, its application would require us to establish control of the rate of convergence of the fourth moment of R^j to $3(\Sigma_t)_{jj}^2$.

First we present a theorem from Rinott (1994). At the expense of stronger conditions (which for us are fulfilled as a result of putting a uniform bound on scaled allelic effects), this result improves the rate of convergence in the corresponding result in Baldi and Rinott (1989). In contrast to the classical result, it allows for sums of non-identically distributed random variables with some local dependence structure which is most conveniently expressed through what is known as a *dependency graph*.

Definition A.1. Let $\{X_l; l \in \mathcal{V}\}$ be a collection of random variables. The graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ denote the vertex set and edge set respectively, is said to be a *dependency graph* for the collection if for any pair of disjoint subsets A_1 and A_2 of $\mathcal{V}$ such that no edge in $\mathcal{E}$ has one endpoint in A_1 and the other in A_2, the sets of random variables $\{X_l; l \in A_1\}$ and $\{X_l; l \in A_2\}$ are independent.

To establish the rate of convergence in Lange (1978) one would take the connected components of the dependency graph to be the chromosomes. The degree of a vertex in the graph is the number of edges connected to it and the maximal degree of the graph is just the maximum of the degrees of the vertices in it.

Theorem A.2 (*Theorem 2.2, Rinott (1994)*). *Let* $Y_1, \ldots, Y_M$ *be random variables having a dependency graph whose maximal degree is strictly less than* D *, satisfying* $|Y_l - \mathbb{E}[Y_l]| \leq B$ *a.s.,* $l = 1, \ldots, M$ *,* $\mathbb{E}[\sum_{l=1}^{M} Y_l] = \lambda$ *and* $\mathrm{Var}\left(\sum_{l=1}^{M} Y_l\right) = \sigma^2 > 0$ *. Then*

$$\left| \mathbb{P}\left[\frac{\sum_{l=1}^{M} Y_l - \lambda}{\sigma} \leq w \right] - \Phi(w) \right|$$

$$\leq \frac{1}{\sigma} \left\{ \sqrt{\frac{1}{2\pi}} DB + 16\left(\frac{M}{\sigma^2}\right)^{1/2} D^{3/2} B^2 + 10\left(\frac{M}{\sigma^2}\right) D^2 B^3 \right\}, \quad (19)$$

where Φ *is the distribution function of a standard normal random variable.*

In particular, when D and B are order one and σ^2 is of order M, as is the case in our applications, this yields a bound of order $1/\sqrt{M}$. Although much of the appeal of this result is that it allows dependence between the variables, we use it in the independent case, when the dependency graph has no edges, and so the maximum degree of any vertex is zero and we may take any $D > 1$.

Rate of convergence in de Jong's CLT

We now turn to a version of the Central Limit Theorem due to de Jong (1990) and its refinements (to include, in particular, the rate of convergence) due to Döbler and Peccati (2017). This result is particularly well-suited to the analysis of epistatic interactions. Consider independent random variables $\chi_1, \chi_2, \ldots, \chi_M$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We write $[M] = \{1, 2, \ldots, M\}$ and for each $I \subseteq [M]$, $\mathcal{F}_I = \sigma(\{X_i : i \in I\})$.

Suppose that $W = f(\chi_1, \ldots, \chi_M)$ satisfies $\mathbb{E}[W] = 0$ and $\mathbb{E}[W^4] < \infty$, then writing

$$W_J = \sum_{L \subseteq J} (-1)^{|J| - |L|} \mathbb{E}[W | \mathcal{F}_L],$$

we have the Hoeffding decomposition

$$W = \sum_{J \subseteq [M]} W_J, \quad (20)$$

in which $\mathbb{E}[W_J] = 0$, $\mathbb{E}[W_J^2] = \sigma_J^2 < \infty$ and $\mathbb{E}[W_I | \mathcal{F}_J] = 0$ unless $I \subseteq J$. In the special case in which $W_J = 0$ unless $|J| = d$, W is called a degenerate U-statistic of order d.

Suppose that W is a degenerate U-statistic of order d, normalised to have variance one. Define

$$\rho(M)^2 = \max_{1 \leq i \leq M} \sum_{K \ni i, |K| = d} \sigma_K^2.$$

Then de Jong's Central Limit Theorem (de Jong, 1990) says that if $\mathbb{E}[W^4] \to 3$ and $\rho(M)^2 \to 0$ as $M \to \infty$ then W converges to a standard normal random variable. Döbler and Peccati (2016) establish the rate of this convergence. They also prove a multidimensional version of the result showing, in particular, that for a vector of degenerate U-statistics based on the same set of random variables $\{\chi_1, \ldots, \chi_M\}$, joint convergence to normality follows from convergence of the marginals. This result is precisely of the right form to extend our results to more complicated epistatic interactions. Since the statement of the multi-dimensional result is rather involved, we satisfy ourselves by stating the one-dimensional result, that is the result applying to a single U-statistic.

The rate of convergence is described in terms of the Kantorovich–Wasserstein distance between the distribution of W and a standard normal distribution. For two probability measures μ and ν on the real line, this is defined as

$$d_{\mathcal{W}}(\mu, \nu) = \sup\left\{ \left| \int h \, d\mu - \int h \, d\nu \right| \right.$$

$$\left. : h \text{ Lipschitz with Lipschitz constant } \|h\|_{\mathrm{Lip}} \leq 1 \right\}.$$

This metric (sometimes called transport distance) is a very natural way to compare the distributions of two random variables when one is derived from the other by small non-uniform perturbations. It is well known, see e.g. Shorack and Wellner (1986) p. 64, that if the cumulative distribution functions corresponding to μ and ν are F_1, F_2 respectively, then

$$d_{\mathcal{W}}(\mu, \nu) = \int_{-\infty}^{\infty} |F_1(t) - F_2(t)| \, dt.$$

Theorem A.3 (*Döbler and Peccati (2017)*, *Theorem 1.3*). *In the setting above, there are universal (and explicit) constants* C_1 *and* C_2 *such that writing* Z *for a standard normal random variable,*

$$d_{\mathcal{W}}(W(M), Z) \leq C_1 \sqrt{|\mathbb{E}[W(M)^4] - 3|} + C_2 \rho(M).$$

This is the analogue of Proposition 3.1 of Chatterjee (2008) for more natural formulations of the epistatic interaction.

Appendix B. Trait distribution in the ancestral population

In this section, we show that as the number of loci tends to infinity, the distribution of the traits $(Z^1, \ldots, Z^{N_0})$ in the ancestral population converges to that of a multivariate normal with mean vector $(\bar{z}_0, \ldots, \bar{z}_0)$ and covariance matrix $\hat{\sigma}^2$ Id. To this end, take $\beta = (\beta_1, \ldots, \beta_{N_0}) \in \mathbb{R}^{N_0}$ and recall our notation $\|\beta\|_1 = \sum_{j=1}^{N_0} |\beta_j|$ and $\|\beta\|_2^2 = \sum_{j=1}^{N_0} \beta_j^2$. We consider $Z_\beta = \sum_{j=1}^{N_0} \beta_j Z^j$.

To apply Theorem A.2, we must first identify the mean and the variance of Z_β. Since we are considering the ancestral population, we have

$$\sum_{j=1}^{N_0} \beta_j Z^j = \bar{z}_0 \sum_{j=1}^{N_0} \beta_j + \frac{1}{\sqrt{M}} \sum_{j=1}^{N_0} \sum_{l=1}^{M} \beta_j \hat{\eta}_l^j.$$

The double sum has mean zero and, since the summands are all independent, variance

$$\mathrm{Var}\left(\frac{1}{\sqrt{M}} \sum_{j=1}^{N_0} \sum_{l=1}^{M} \beta_j \hat{\eta}_l^j \right) = \frac{1}{M} \sum_{j=1}^{N_0} \sum_{l=1}^{M} \beta_j^2 \mathrm{Var}(\hat{\eta}_l) = \|\beta\|_2^2 \hat{\sigma}_M^2.$$

We shall apply Theorem A.2 to the quantities $Y_l = \sum_{j=1}^{N_0} \beta_j \hat{\eta}_l^j$. Since they are independent, we can take $D = 2$ and since, by assumption, the scaled allelic effects are bounded by B, $|\sum_{j=1}^{N_0} \beta_j \hat{\eta}_{jl}| \leq B\|\beta\|_1$ for all l. Then,

$$\left| \mathbb{P}\left[\frac{\sum_{j=1}^{N_0} \beta_j (Z^j - \bar{z}_0)}{\|\beta\|_2 \hat{\sigma}_M} \leq z \right] - \Phi(z) \right|$$

$$\leq \frac{C}{\|\beta\|_2 \sqrt{M} \hat{\sigma}_M} \left(1 + \frac{1}{\|\beta\|_2^2 \hat{\sigma}_M^2} \right), \quad (21)$$

where Φ is the cumulative distribution function of a standard normal random variable and the constant C has an explicit expression (depending on B and $\|\beta\|_1$), which can be read off from Theorem A.2. In particular, in the special case when $\beta_j = 0$ for all $j \neq k$ and $\beta_k = 1$, so that $\sum_{j=1}^{N_0} \beta_j Z^j = Z^k$, the genetic component of the trait of the kth individual, the constant C is independent of N_0.

Since the vector β is arbitrary, this proves convergence of the joint distribution of traits in the ancestral population to a multivariate normal with the given mean vector and covariance matrix.

Appendix C. Observed traits and scaled allelic effects in the ancestral population

When we condition on the observed trait values in our pedigree, we gain some information on the scaled allelic effect at each locus. In order to control the magnitude of this effect we use Bayes' rule to turn it into a question about the effect of knowing the allelic effect at a given locus on the probability of observing a particular trait. We then need to be able to control

$$
\frac{\mathbb{P}\left[\widetilde{Z}^{j[1]} - \frac{1}{\sqrt{M}}\eta_l^{j[1]} = z_1 - \frac{x}{\sqrt{M}}\right]}{\mathbb{P}\left[\widetilde{Z}^{j}[1] = z_1\right]},
$$

where we are interpreting the probabilities as density functions.

We begin with the case in which the parents are from the ancestral population. Using the result of Appendix B, the observed trait in each individual is, up to an error of order $1/\sqrt{M}$, (independently) distributed according to the sum of a normal random variable with mean $\bar{z}_0$ and variance $\widehat{\sigma}_M^2$ and an independent normal with mean zero and variance σ_E^2.

Let us write $p(\sigma^2, \mu, y)$ for the density function of a normally distributed random variable with mean μ and variance σ^2. Then, taking $\beta_k = 0$ for $k \neq j$ and $\beta_j = 1$ gives

$$
\left|\mathbb{P}\left[\widetilde{Z}^{j[1]} = z_1\right] - p(\sigma_E^2 + \widehat{\sigma}_M^2, \bar{z}_0, z_1)\right| \leq \frac{C}{\sqrt{M}\widehat{\sigma}_M}\left(1 + \frac{1}{\widehat{\sigma}_M^2}\right).
$$

Remark C.1. The Central Limit Theorem A.2 of Appendix B only gives convergence of the cumulative distribution functions of the genetic component of the ancestral traits with a rate $1/\sqrt{M}$. If there is a differentiable density function for each M then we can deduce the same order of convergence for the density function. If, for example, allelic effects are discrete, then additional conditions would be required to approximate this ratio of probabilities by the corresponding normal distribution with this degree of accuracy as we need a *local* limit theorem to hold. McDonald (2005) surveys results in this direction. Without such conditions, the rate of convergence can be shown to be at least $1/M^{1/4}$, but simple counterexamples show that this is optimal. Convolution with the environmental noise rescues us and gives the faster rate of convergence of densities reported here.

The same result applied to $\widetilde{Z}^{j[1]} - \eta_l^j/\sqrt{M}$, gives, up to an error of order $1/M$ which we ignore,

$$
\begin{aligned}
&\left|\mathbb{P}\left[\widetilde{Z}^{j[1]} - \frac{1}{\sqrt{M}}\eta_l^j[1] = z_1 - \frac{x}{\sqrt{M}}\right]\right.\\
&\quad \left. - p\left(\sigma_E^2 + \widehat{\sigma}_M^2, \bar{z}_0, z_1 - \frac{x}{\sqrt{M}}\right)\right|\\
&\leq \frac{C'}{\sqrt{M}\widehat{\sigma}_M}\left(1 + \frac{1}{\widehat{\sigma}_M^2}\right),
\end{aligned}
$$

where we recall that $\bar{z}_0$ is the mean of the genetic component of the trait in generation zero. Performing a Taylor expansion of $p(\sigma_E^2 + \widehat{\sigma}_M^2, \bar{z}_0, z_1 - x/\sqrt{M})$ around $z_1 - \bar{z}_0$ and using that

$$
\frac{p'(\sigma^2, 0, y)}{p(\sigma^2, 0, y)} \leq \frac{|y|}{\sigma^2},
$$

we see that for parents of individuals in the first generation,

$$
\left|\frac{\mathbb{P}\left[\widetilde{Z}^{j[1]} - \frac{1}{\sqrt{M}}\eta_l^{j[1]} = z_1 - \frac{x}{\sqrt{M}}\right]}{\mathbb{P}\left[\widetilde{Z}^{j[1]} = z_1\right]} - 1\right| \leq \frac{1}{\sqrt{M}}C\left(\widehat{\sigma}_M^2, |z_1 - \bar{z}_0|\right),
$$

where $C(\sigma^2, |z|)$ was defined in Eq. (12). Just as with our toy model at the beginning of Section 3, we see that the approximation requires that the trait we are sampling is not 'too extreme'.

Appendix D. One generation of reproduction

In order to include (house of cards) mutation, we introduce another collection of Bernoulli random variables. We write $\mathcal{M}_l^j = 1$ if there is a mutation at locus l in individual j; otherwise it is zero. Under our assumption of a constant probability of mutation across all loci, we have $\mathbb{P}[\mathcal{M}_l^j = 1] = \mu = 1 - \mathbb{P}[\mathcal{M}_l^j = 0]$.

We now establish that after one round of mating, conditional on knowing $\mathcal{P}(1)$ and $\widetilde{Z}(0) = \underline{z}$,

$$
\left(Z^j - \mu\bar{z}_\mu - (1-\mu)\frac{Z^{j[1]} + Z^{j[2]}}{2}\right)_{j=1,\dots,N_1}
$$

converges in distribution to a mean zero multivariate normally distributed random variable with diagonal variance–covariance matrix Σ_1^μ with on-diagonal entries $(\Sigma_1^\mu)_{jj}$ given by the limit of (13) (or rather the full version, Eq. (24), which includes terms of order μ^2).

The 'remainder term' R^j in (9) is given by

$$
\begin{aligned}
R^j = {} & \frac{1}{\sqrt{M}}\sum_{l=1}^{M}\left(\mathcal{M}_l^j\check{\eta}_l^j - \mu\mathbb{E}[\check{\eta}_l]\right)\\
& + \frac{1}{\sqrt{M}}\sum_{l=1}^{M}\left((1-\mathcal{M}_l^j)X_l^j - \frac{1-\mu}{2}\right)\eta_l^{j[1]}\\
& + \frac{1}{\sqrt{M}}\sum_{l=1}^{M}\left((1-\mathcal{M}_l^j)(1-X_l^j) - \frac{1-\mu}{2}\right)\eta_l^{j[2]}. \quad (22)
\end{aligned}
$$

The Bernoulli random variables $\mathcal{M}_l^j$ and X_l^j are independent of both $\mathcal{P}(1)$ and $\widetilde{Z}(0)$ and so $\mathbb{E}[R^j|\mathcal{P}(1), \widetilde{Z}(0) = \underline{z}] = 0$. Moreover, since the Bernoulli variables in different individuals are independent, for $i \neq j$, $\mathbb{E}[R^iR^j|\mathcal{P}(1), \widetilde{Z}(0) = \underline{z}] = 0$. To establish $\mathbb{E}[(R^j)^2|\mathcal{P}(1), \widetilde{Z}(0) = \underline{z}]$, first we use Bayes' rule to control the conditional distribution of $\eta_l^{j[1]}$. We condition on the whole vector of observed traits $\widetilde{Z}(0) = \underline{z}$, but since individuals in our ancestral population are assumed unrelated, from the perspective of $\eta_l^{j[1]}$, this is equivalent to conditioning on the observed trait $\widetilde{Z}^{j[1]}$ of the first parent of the j th individual. It is convenient to write z_1 for the corresponding coordinate of $\underline{z}$.

$$
\begin{aligned}
&\mathbb{P}\left[\eta_l^{j[1]} = x\big|\mathcal{P}(1), \widetilde{Z}(0) = \underline{z}\right] = \mathbb{P}\left[\eta_l^{j[1]} = x\big|\widetilde{Z}^{j[1]} = z_1\right]\\
&= \frac{\mathbb{P}\left[\eta_l^{j[1]} = x, \widetilde{Z}^{j[1]} - \frac{x}{\sqrt{M}} = z_1 - \frac{x}{\sqrt{M}}\right]}{\mathbb{P}\left[\widetilde{Z}^{j[1]} = z_1\right]}\\
&= \mathbb{P}\left[\eta_l^{j[1]} = x\right]\frac{\mathbb{P}\left[\widetilde{Z}^{j[1]} - \frac{1}{\sqrt{M}}\eta_l^{j[1]} = z_1 - \frac{x}{\sqrt{M}}\right]}{\mathbb{P}\left[\widetilde{Z}^{j[1]} = z_1\right]},
\end{aligned}
$$

where we have used independence of inheritance at different loci and the ratio on the right should be interpreted as a ratio of probability density functions. We showed in Appendix C that the ratio in the last line differs from one by at most

$$
\frac{1}{\sqrt{M}}C(\widehat{\sigma}_M^2, |z_1 - \bar{z}_0|).
$$

Since individuals in the ancestral population are assumed to be unrelated, $\eta_l^{j[1]}$ and $\eta_l^{j[2]}$ are independent and so combining the calculation above with the symmetric one for $\eta_l^{j[2]}$ we can calculate that for some $\alpha \in [-1, 1]$,

$$
\begin{aligned}
\Sigma_1^{M,\mu} &:= \mathbb{E}\left[(R^j)^2\,\big|\,\mathcal{P}(1), \widetilde{Z}(0) = \underline{z}\right]\\
&= \mathbb{E}\left[(R^j)^2\,\big|\,\mathcal{P}(1)\right]\left(1 + \frac{\alpha}{\sqrt{M}}C(\widehat{\sigma}_M^2, |z_1 - \bar{z}_0|)\right).
\end{aligned}
$$

Noting that inheritance at different loci is independent, the variance of R^j will be the sum of the variances at each locus. We

consider the summand corresponding to a single locus, l say. Omitting the factor of $1/M$, the square of the first term, corresponding to mutation, contributes

$$\mu(1-\mu)\mathbb{E}[\check{\eta}_l]^2 + \mu\mathrm{Var}(\check{\eta}_l).$$

Since the variances of $(1-\mathcal{M}_l^j)X_l^j$ and $(1-\mathcal{M}_l^j)(1-X_l^j)$ are both $(1-\mu)(1+\mu)/4$, the squares of the next two terms contribute

$$\frac{(1-\mu^2)}{4}\mathbb{E}[(\eta_l^{j[1]})^2 + (\eta_l^{j[2]})^2].$$

The cross terms are also non-trivial.

$$2\mathbb{E}\left[(\mathcal{M}_l^j\check{\eta}_l - \mu\mathbb{E}[\check{\eta}_l])\left((1-\mathcal{M}_l^j)X_l^j - \frac{1-\mu}{2}\eta_l^{j[1]}\right)\right]$$
$$= -\mu(1-\mu)\mathbb{E}[\check{\eta}_l]\mathbb{E}[\eta_l^{j[1]}].$$

Similarly,

$$2\mathbb{E}\left[(\mathcal{M}_l^j\check{\eta}_l - \mu\mathbb{E}[\check{\eta}_l])\left((1-\mathcal{M}_l^j)(1-X_l^j) - \frac{1-\mu}{2}\eta_l^{j[2]}\right)\right]$$
$$= -\mu(1-\mu)\mathbb{E}[\check{\eta}_l]\mathbb{E}[\eta_l^{j[2]}].$$

Finally,

$$2\mathbb{E}\left[\left((1-\mathcal{M}_l^j)X_l^j - \frac{1-\mu}{2}\right)\eta_l^{j[1]}\left((1-\mathcal{M}_l^j)(1-X_l^j) - \frac{1-\mu}{2}\right)\eta_l^{j[2]}\right]$$
$$= -\frac{(1-\mu)^2}{2}\mathbb{E}[\eta_l^{j[1]}\eta_l^{j[2]}].$$

Combining these, we obtain

$$\mu(1-\mu)\mathbb{E}[\check{\eta}_l]^2 + \mu\mathrm{Var}(\check{\eta}_l) + \frac{(1-\mu)^2}{4}\mathbb{E}[(\eta_l^{j[1]} - \eta_l^{j[2]})^2]$$
$$+ \mu(1-\mu)\left(\mathbb{E}[(\eta_l^{j[1]})^2] - 2\mathbb{E}[\check{\eta}_l]\mathbb{E}[\eta_l^{j[1]}] + \mathbb{E}[\check{\eta}_l^2]\right)$$
$$- \mu(1-\mu)\mathbb{E}[\check{\eta}_l^2]$$
$$= \mu^2\mathrm{Var}(\check{\eta}_l) + \frac{(1-\mu)^2}{4}\mathbb{E}[(\eta_l^{j[1]} - \eta_l^{j[2]})^2]$$
$$+ \mu(1-\mu)\mathbb{E}[(\eta_l^{j[1]} - \check{\eta}_l)^2]$$

and, since $\eta_l^{j[1]}$ is sampled from the ancestral population and so is a copy of $\widehat{\eta}_l$, this yields

$$\mu^2\mathrm{Var}(\check{\eta}_l) + \frac{(1-\mu)^2}{4}\mathbb{E}[(\eta_l^{j[1]} - \eta_l^{j[2]})^2] + \mu(1-\mu)\mathbb{E}[(\widehat{\eta}_l - \check{\eta}_l)^2].$$

(Note that if the individual was produced by selfing, the second term is 0.) It is immediate from this calculation that the variance of our limiting distribution of traits is Σ_1^μ, as claimed. To check that the limit is a multivariate normal, we mimic what we did in the ancestral population: for an arbitrary vector $\beta = (\beta_1, \ldots, \beta_{N_1})$ we show that $\sum_{j=1}^{N_1}\beta_j R^j$ converges to a normal random variable as $M \to \infty$. As before the strategy is to apply Theorem A.2. This time

$$\sum_{j=1}^{N_1}\beta_{jR}^j = \frac{1}{\sqrt{M}}\sum_{l=1}^{M}Y_l,$$

where

$$Y_l = \sum_{j=1}^{N_1}\beta_j\Bigg\{\mathcal{M}_l^j\check{\eta}_l^j - \mu\mathbb{E}[\check{\eta}_l] + \left((1-\mathcal{M}_l^j)X_l^j - \frac{1-\mu}{2}\right)\eta_l^{j[1]}$$
$$+ \left((1-\mathcal{M}_l^j)(1-X_l^j) - \frac{1-\mu}{2}\right)\eta_l^{j[2]}\Bigg\}. \tag{23}$$

Each such term is bounded by $B\|\beta\|_1$ and inheritance is independent at distinct loci and so Theorem A.2 yields convergence (in law) of $\sum_{l=1}^{M}Y_l/\sqrt{M} = \sum_{j=1}^{N_1}\beta_j R^j$ to a mean zero normal random

variable with variance

$$\sum_{j=1}^{N_1}\beta_j^2(\Sigma_1^\mu)_{jj},$$

from which, since β was arbitrary, we deduce convergence of $(R^1, \ldots, R^{N_1})$, conditional on knowing $\mathcal{P}(1)$ (the parents of each individual in the population) and $\widetilde{Z}(0)$ (the observed traits of all parents) to a multivariate normal with mean zero and diagonal variance–covariance matrix with on-diagonal entries identically equal to Σ_1. More precisely, just as in (21),

$$\left|\mathbb{P}\left[\frac{\sum_{j=1}^{N_1}\beta_j R^j}{\sqrt{\sum_{j=1}^{N_1}\beta_j^2(\Sigma_1^{M,\mu})_{jj}}} \le y \Big| \mathcal{P}(1), \widetilde{Z}(0) = \underline{z}\right] - \Phi(y)\right|$$
$$\le \frac{C'}{\sqrt{M\sum_{j=1}^{N_1}\beta_j^2(\Sigma_1^{M,\mu})_{jj}}}\left(1 + \frac{1}{\sum_{j=1}^{N_1}\beta_j^2(\Sigma_1^{M,\mu})_{jj}}\right)$$
$$\times \left(1 + C(\widehat{\sigma}_M^2, |z_1 - \bar{z}_0|)\right).$$

Appendix E. Generation t

We now provide the missing steps in the general case. We proceed by induction.

Suppose that we have proved the asymptotic normality of the vector of genetic components of trait values and (15) (that conditioning on the pedigree and the observed ancestral traits provides negligible information about the distribution of allelic types at a given locus) for all generations up to and including $(t-1)$. We have already checked generation one.

We first prove (15). Let us write $A \sim \eta_l^{j[1]}$ to mean that A is the set of individuals in $\mathcal{P}(t-1)$ that are identical by descent at locus l with the first parent of individual j in generation t. Note that A depends on the pedigree and the Bernoulli random variables that determine inheritance at the lth locus, but not on the value $\eta_l^{j[1]}$ and so partitioning on the set A,

$$\mathbb{P}\left[\widetilde{Z}(t-1) = \underline{z} | \mathcal{P}(t), \eta_l^{j[1]} = x\right]$$
$$= \sum_A \mathbb{P}\left[\widetilde{Z}(t-1) = \underline{z} | A \sim \eta_l^{j[1]}, \mathcal{P}(t), \eta_l^{j[1]} = x\right]$$
$$\times \mathbb{P}\left[A \sim \eta_l^{j[1]} | \mathcal{P}(t)\right].$$

We write a for the eldest individual in A and $|a|$ for the generation in which it lived. Evidently the trait values in $\mathcal{P}(t-1) \setminus A$ do not depend on $\eta_l^{j[1]}$. Moreover, if we further partition on the value of Z^a (the genetic component of the trait of the eldest member of A), we see that for all $a' \in A \setminus a$, the probability that $\widetilde{Z}^{a'} - z_a = z_{a'} - z_a$ is independent of the value of $\eta_l^{j[1]}$. In other words, the dependence of the trait values in the pedigree on $\eta_l^{j[1]}$ is entirely captured by

$$\mathbb{P}\left[\widetilde{Z}^a = z_a | \widetilde{Z}(|a| - 1), \mathcal{P}(|a|), \eta_l^a = x\right].$$

Since a lives at the latest in generation $t-1$, we can use our inductive hypothesis to write that

$$\mathbb{P}\left[\widetilde{Z}^a = z_a | \widetilde{Z}(|a| - 1) = \underline{z}(|a| - 1), \mathcal{P}(t), \eta_l^a = x\right]$$
$$= \mathbb{P}\left[\widetilde{Z}^a = z_a | \widetilde{Z}(|a| - 1) = \underline{z}(|a| - 1), \mathcal{P}(t)\right]$$
$$\times \left(1 + \frac{\alpha}{\sqrt{M}}C(\underline{\Sigma}_t^M, \bar{\Delta}_t^M(\underline{z}))\right).$$

Since we have successfully eliminated all the conditioning on the value of $\eta_l^{j[1]}$, we can now rearrange our calculations to give Eq. (15) and

$$\mathbb{P}\left[\eta_l^{j[1]} = x \big| \mathcal{P}(t), \widetilde{Z}(t-1) = \underline{z}\right] = \mathbb{P}\left[\eta_l^{j[1]} = x \big| \mathcal{P}(t)\right]$$
$$\times \left(1 + \frac{\alpha}{\sqrt{M}} C(\underline{\Sigma}_t^M, \bar{\Delta}_t^M(\underline{z}))\right).$$

We can perform entirely analogous calculations for the joint law of $\eta_l^{j[1]}$ and $\eta_l^{j[2]}$.

Now consider the mean zero random variable R^j. That $\mathbb{E}[R^i R^j | \mathcal{P}(t), \widetilde{Z}(t-1)] = 0$ for $i \neq j$ follows exactly as before and the calculation of $\mathbb{E}[(R^i)^2 | \mathcal{P}(t), \widetilde{Z}(t-1)]$ also proceeds almost exactly as for generation one. The only distinction is that

$$\mathbb{E}[\eta_l^{j[1]} | \mathcal{P}(t)] = (1-\mu)^{t-1} \mathbb{E}[\widehat{\eta}_l] + (1 - (1-\mu)^{t-1}) \mathbb{E}[\check{\eta}_l],$$

and similarly

$$\mathbb{E}[(\eta_l^{j[1]})^2 | \mathcal{P}(t)] = (1-\mu)^{t-1} \mathbb{E}[\widehat{\eta}_l^2] + (1 - (1-\mu)^{t-1}) \mathbb{E}[\check{\eta}_l^2],$$

and so the contribution to $\mathbb{E}[(R^i)^2 | \mathcal{P}(t), \widetilde{Z}(t-1)]$ from the l th locus becomes

$$(1 - (1-\mu)^2)\mathrm{Var}(\check{\eta}_l) + \frac{(1-\mu)^2}{4}\mathbb{E}[(\eta_l^{j[1]} - \eta_l^{j[2]})^2 | \mathcal{P}(t)]$$
$$+ 2\mu(1-\mu)^t \left(\frac{\mathbb{E}[(\widehat{\eta}_l - \check{\eta}_l)^2]}{2} - \mathrm{Var}(\check{\eta}_l)\right).$$

Summing over loci yields

$$(\Sigma_t^{M,\mu})_{jj} = \frac{(1-\mu)^2}{4} \frac{1}{M} \sum_{l=1}^M \mathbb{E}\left[\left(\eta_l^{j[1]} - \eta_l^{j[2]}\right)^2 \Big| \mathcal{P}(t)\right]$$
$$+ (1 - (1-\mu)^2)\check{\sigma}_M^2$$
$$+ 2\mu(1-\mu)^t \frac{1}{M}\sum_{l=1}^M \left(\frac{\mathbb{E}[(\widehat{\eta}_l - \check{\eta}_l)^2]}{2} - \mathrm{Var}(\check{\eta}_l)\right), \quad (24)$$

which for small μ becomes (13).

Now, exactly as we did for generation one, we can fix $\beta_1, \ldots, \beta_{N_t} \in \mathbb{R}$ and apply Theorem A.2 to Y_l given by (23) (with the same bound) to deduce that conditional on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1) = \underline{z}$,

$$\left(R^1, \ldots, R^{N_t}\right) \to \mathcal{N}(0, \Sigma_t^\mu).$$

In particular,

$$\left| \mathbb{P}\left[\frac{Z^j - \mu \bar{z}_\mu - (1-\mu)\frac{Z^{j[1]} + Z^{j[2]}}{2}}{\sqrt{(\Sigma_t^{M,\mu})_{jj}}} \leq y \Big| \mathcal{P}(t), \widetilde{Z}(t-1) = \underline{z}\right] - \Phi(y)\right|$$
$$\leq \frac{1}{\sqrt{M}} C(\underline{\Sigma}_t^M, \bar{\Delta}_t^{M,\mu}(\underline{z})). \quad (25)$$

Appendix F. Environmental noise: conditioning multivariate Gaussian vectors

In order to estimate the proportion of an observed trait that is due to environmental noise, and thus make predictions about offspring traits, we need a standard result for conditioning multivariate normal random vectors on their marginal values which, for ease of reference, we record here.

Theorem F.1. *Suppose that*
$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}\right).$$
Then
$$x_A | x_B \sim \mathcal{N}\left(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(x_B - \mu_B), \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}\right).$$

The proof can be found e.g. in Brockwell and Davis (1996) (Prop. 1.3.1 in Appendix A). We write $E_c(t) = (E_c^j(t))_{j=1,\ldots,N_t}$ for

the conditioned vector $(E^1, \ldots, E^{N_t}) | \mathcal{P}(t), \widetilde{Z}(t)$. To see how Theorem F.1 leads to a recurrence for the mean and variance of $E_c(t)$, we begin with generation zero. In this case there are just two components to consider, $(R^j)_{j=1,\ldots,N_0}$ and $(E^j)_{j=1,\ldots,N_0}$, each of which is (at least asymptotically) a mean zero Gaussian with diagonal variance–covariance matrix. We wish to calculate $x_A | x_B$ where $x_A = (E^j)_{j=1,\ldots,N_0}$ and $x_B = (\widetilde{Z}^j - \bar{z}_0)_{j=1,\ldots,N_0}$. We have

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix}\right)$$

where $\Sigma_{AA} = \sigma_E^2 \mathrm{Id}$, $\Sigma_{BB} = (\widehat{\sigma}^2 + \sigma_E^2)\mathrm{Id}$ and $\Sigma_{AB} = \Sigma_{BA} = \sigma_E^2 \mathrm{Id}$. Applying Theorem F.1, $(E_c^j(0))_{j=1,\ldots,N_0}$ (as $M \to \infty$) is a Gaussian random variable with mean vector

$$A_c(0) = \left(\frac{\sigma_E^2}{\widehat{\sigma}^2 + \sigma_E^2}(\widetilde{Z}_j - \bar{z}_0)\right)_{j=1,\ldots,N_0}$$

and variance–covariance matrix

$$\Sigma_c^E(0) = \frac{\widehat{\sigma}^2 \sigma_E^2}{\widehat{\sigma}^2 + \sigma_E^2}\mathrm{Id}.$$

For the recursive step, we now set

$$x_A = (E^j)_{j=1,\ldots,N_t}, \quad x_B = (\Delta\widetilde{Z}^j(t))_{j=1,\ldots,N_t}.$$

Then, writing $\mathbb{E}_c$ to indicate that we are conditioning on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1)$,

$$\mathbb{E}_c[x_A] = (0, \ldots, 0), \quad \mathbb{E}_c[x_B]$$
$$= \left(-(1-\mu)\frac{A^{j[1]}(t-1) + A^{j[2]}(t-1)}{2}\right)_{j=1,\ldots,N_t},$$

$$\Sigma_{AA} = \sigma_E^2 \mathrm{Id}, \quad \Sigma_{AB} = \Sigma_{BA} = \sigma_E^2 \mathrm{Id}.$$

The more complex term is

$$\Sigma_{BB} = \left(\mathrm{Cov}(\Delta\widetilde{Z}^i(t), \Delta\widetilde{Z}^j(t)) | \mathcal{P}(t), \widetilde{Z}(t-1)\right)_{i,j=1,\ldots,N_t}.$$

We treat the cases $i = j$ and $i \neq j$ separately. In the expression below, $i[a]$ is the a th parent of individual i (with $a \in \{1, 2\}$). First suppose that $i \neq j$, again conditionally on $\mathcal{P}(t)$ and $\widetilde{Z}(t-1)$,

$$\mathbb{E}_c[\Delta\widetilde{Z}^i(t)\Delta\widetilde{Z}^j(t)] - \mathbb{E}_c[\Delta\widetilde{Z}^i(t)]\mathbb{E}_c[\Delta\widetilde{Z}^j(t)]$$
$$= (1-\mu)^2 \mathbb{E}_c\left[\frac{(E^{i[1]}(t) + E^{i[2]}(t))}{2}\frac{(E^{j[1]}(t) + E^{j[2]}(t))}{2}\right]$$
$$- (1-\mu)^2 \mathbb{E}_c\left[\frac{E^{i[1]}(t) + E^{i[2]}(t)}{2}\right]\mathbb{E}_c\left[\frac{E^{j[1]}(t) + E^{j[2]}(t)}{2}\right]$$
$$= \frac{(1-\mu)^2}{4}\sum_{a,b\in\{1,2\}}\mathrm{Cov}(E_c^{i[a]}(t), E_c^{i[b]}(t))$$
$$= \frac{(1-\mu)^2}{4}\sum_{a,b\in\{1,2\}}(\Sigma_c^E(t-1))_{i[a],j[b]}.$$

If $i = j$,

$$\mathbb{E}_c\left[\left(\Delta\widetilde{Z}^i(t)\right)^2\right] - \mathbb{E}_c\left[\Delta\widetilde{Z}^i(t)\right]^2 = \sigma_E^2 + (\Sigma_{t-1})_{ii}$$
$$+ (1-\mu)^2 \mathbb{E}_c\left[\left(\frac{E^{i[1]}(t) + E^{i[2]}(t)}{2}\right)^2\right]$$
$$- (1-\mu)^2 \mathbb{E}_c\left[\frac{E^{i[1]}(t) + E^{i[2]}(t)}{2}\right]^2$$
$$= \sigma_E^2 + (\Sigma_{t-1})_{ii} + \frac{(1-\mu)^2}{4}\sum_{a,b\in\{1,2\}}\left(\Sigma_c^E(t-1)\right)_{i[a],i[b]}.$$

Again applying Theorem F.1, we obtain that $E_c(t)$ has mean vector

$$A_c(t) = \left(\sigma_E^2 \Sigma_{BB}^{-1}\left(\Delta\widetilde{z}_j^t + (1-\mu)\frac{A^{j[1]}(t-1) + A^{j[2]}(t-1)}{2}\right)\right)_{j=1,\ldots,N_t},$$

and variance–covariance matrix

$$\Sigma_c^E(t) = \sigma_E^2 \mathrm{Id} - \sigma_E^4 \Sigma_{BB}^{-1}.$$

Appendix G. A coupling argument

We do not spell out all of the details of the proof of convergence to a multivariate normal in the presence of epistasis. However, we illustrate a useful coupling argument by explaining how to use it to prove that in generation one, $\mathbb{E}[\Delta Z^j]$ is of order $1/\sqrt{M}$. Recall that

$$\Delta Z^j = \frac{1}{\sqrt{M}} \sum_{l=1}^{M} \left(\eta_l^j - \frac{\eta_l^{j[1]} + \eta_l^{j[2]}}{2} \right)$$
$$+ \frac{1}{M} \sum_{1 \le l \le m \le M} \left(\eta_{lm} - \frac{\eta_{lm}^{j[1]} + \eta_{lm}^{j[2]}}{4} \right).$$

We use the notation $\mathbf{1}_{U \leftarrow j[1]}$ for the Bernoulli random variable which takes the value 1 when all the loci in U were inherited from $j[1]$. Then

$$\Delta Z^j = \frac{1}{\sqrt{M}} \sum_{l=1}^{M} \left(\eta_l^j - \frac{\eta_l^{j[1]} + \eta_l^{j[2]}}{2} \right)$$
$$+ \frac{1}{M} \sum_{1 \le l \le m \le M} \left(\eta_{lm}^{j[1]} \left(\mathbf{1}_{\{l,m\} \leftarrow j[1]} - \frac{1}{4} \right) + \eta_{lm}^{j[2]} \left(\mathbf{1}_{\{l,m\} \leftarrow j[2]} - \frac{1}{4} \right) \right)$$
$$+ \frac{1}{M} \sum_{1 \le l \le m \le M} \eta_{lm}^j \left(1 - \mathbf{1}_{\{l,m\} \leftarrow j[1]} - \mathbf{1}_{\{l,m\} \leftarrow j[2]} \right).$$

Because the Bernoulli random variables that determine inheritance are independent of the parental allelic effects, the expectation of the first sum is zero. A priori, the second term could be order one, but we now argue that it is order $1/\sqrt{M}$. The idea is a simple coupling argument. The analogue of (15) tells us that even conditioned on the values z_A, $z_{A(2)}$ in the ancestral population, for any fixed pair of loci $\{l, m\}$, the allelic state of the parents $j[1]$, $j[2]$ at those loci have the original distribution $\widehat{\chi}$ with probability $1 - \mathcal{O}(1/\sqrt{M})$. We can couple the conditioned distributions of the allelic states for the loci $\{l, m\}$ in each parent in such a way that (independently at the two loci), with probability $1 - C/\sqrt{M}$, χ_l (resp. χ_m) is drawn from the unbiased distribution $\widehat{\chi}_l$ (resp. $\widehat{\chi}_m$) and with probability $C/\sqrt{M}$, χ_l (resp. χ_m) is drawn from some modified distribution χ_l^* (resp. χ_m^*), which is independent of $\widehat{\chi}_l$, $\widehat{\chi}_m$. We now sum over all admissible inheritance patterns.

$$\mathbb{E}\left[\eta_{lm}^j \left(1 - \mathbf{1}_{\{l,m\} \leftarrow j[1]} - \mathbf{1}_{\{l,m\} \leftarrow j[2]} \right) \right]$$
$$= \frac{1}{4} \mathbb{E}\left[\phi_{lm}(\chi_l^{j[1]}, \chi_m^{j[2]}) \right] + \frac{1}{4} \mathbb{E}\left[\phi_{lm}(\chi_m^{j[1]}, \chi_l^{j[2]}) \right]$$
$$+ \frac{1}{4} \left\{ \left(1 - \frac{C}{\sqrt{M}} \right)^2 \mathbb{E}[\phi_{lm}(\widehat{\chi}_l, \widehat{\chi}_m)] \right.$$
$$+ \frac{C}{\sqrt{M}} \left(1 - \frac{C}{\sqrt{M}} \right) \mathbb{E}[\phi_{lm}(\widehat{\chi}_l, \chi_m^*)]$$
$$\left. + \frac{C}{\sqrt{M}} \left(1 - \frac{C}{\sqrt{M}} \right) \mathbb{E}[\phi_{lm}(\chi_l^*, \widehat{\chi}_m)] + \frac{C^2}{M} \mathbb{E}[\phi_{lm}(\chi_l^*, \chi_m^*)] \right\}.$$

The first term is zero by assumption and (18) guarantees that so are the second and third terms. We must multiply the final term by $1/M$ and sum over all loci. The uniform bound on $\sum_{m=1}^{M} \eta_{lm}/\sqrt{M}$ is enough to guarantee that the result is a term of order at most $1/\sqrt{M}$.

References

Abney, M., McPeek, M.S., Ober, C., 2000. Estimation of variance components of quantitative traits in inbred populations. Am. J. Hum. Genet. 66, 629–650.

Baldi, P., Rinott, Y., 1989. On normal approximations of distributions in terms of dependency graphs. Ann. Probab. 17, 1646–1650.

Barton, N.H., 2010. What role does natural selection play in speciation? Philos. Trans. R. Soc. London [Biol] 365, 1825–1840.

Barton, N.H., Etheridge, A.M., 2011. The relation between reproductive value and genetic contribution. Genetics 188, 953–973.

Barton, N.H., Keightley, P.D., 2002. Understanding quantitative genetic variation. Nature Rev. Genet. 3 (1), 11–21.

Barton, N.H., Shpak, M., 2000. The stability of symmetrical solutions to polygenic models. Theor. Popul. Biol. 57, 249–264.

Barton, N.H., Turelli, M., 2004. Effects of genetic drift on variance components under a general model of epistasis. Evolution 58, 2111–2132.

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell 169, 1177–1186.

Brockwell, P.J., Davis, R.A., 1996. Introduction to Time Series and Forecasting. Springer texts in statistics.

Bulmer, M.G., 1971. The effect of selection on genetic variability. Am. Nat. 105, 201–211.

Bulmer, M.G., 1974. Linkage disequilibrium and genetic variability. Genet. Res. 23, 281–289.

Bulmer, M.G., 1980. The Mathematical Theory of Quantitative Genetics. Oxford University Press.

Bulmer, M.G., 1998. Galton's law of ancestral heredity. Heredity 81, 579–585.

Bulmer, M.G., 2004. Did Jenkin's swamping argument invalidate Darwin's theory of natural selection? Br. J. Hist. Sci. 37, 1–17.

Bürger, R., 2000. The Mathematical Theory of Selection, Recombination, and Mutation. Wiley, Chichester.

Cavalli-Sforza, L.L., Bodmer, W.F., 1971. The Genetics of Human Populations. Freeman, San Francisco.

Cavalli-Sforza, L.L., Feldman, M.W., 1976. Evolution of continuous variation: direct approach through joint distribution of genotypes and phenotypes. Proc. Natl. Acad. Sci. USA 73, 1689–1692.

Charlesworth, B., 2015. Causes of natural variation in fitness: Evidence from studies of Drosophila populations. Proc. Natl. Acad. Sci. USA 112 (6), 1662–1669.

Chatterjee, S., 2008. A new method of normal approximation. Ann. Probab. 36 (4), 1584–1610.

Cockerham, C.C., 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39, 859–882.

Cooley, J.W., Tukey, J.W., 1965. An algorithm for the machine calculation of complex Fourier series. Math. Comp. 19, 297–301.

Cotterman, C.W., 1940. A Calculus for Statistico-Genetics (Ph.D. thesis).

Crow, J.F., Felsenstein, J., 1968. The effect of assortative mating on the genetic composition of a population. Eugenics Quart. 15, 85–97.

Davis, A.S., 1871. The north British Review and the Origin of Species. Nature 5, 161.

Dawson, K.J., 1997. Linkage disequilibrium and the infinitesimal limit. Theor. Popul. Biol. 52, 1–18.

de Jong, P., 1990. A central limit theorem for generalized multilinear forms. J. Multivariate Anal. 34, 275–289.

Diehl, S.R., Bush, G.L., 1989. The role of habitat preference in adaptation and speciation. In: Speciation and Its Consequences. Sinauer Press, Sunderland Mass, pp. 345–365.

Döbler, C., Peccati, G., 2017. Quantitative de Jong theorems in any dimension. Electron. J. Probab. 22 (2), 1–35.

Doebeli, M., 1996. A quantitative genetic competition model for sympatric speciation. J. Evol. Biol. 9, 893–910.

Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Proc. Roy. Soc. Edinburgh 52, 399–433.

Frankham, R., 1996. Relationship of genetic variation to population size in wildlife. Conserv. Biol. 10, 1500–1508.

Galton, F., 1877. Typical laws of heredity. Nature 15, 492–495, 512–514, 532–533.

Galton, F., 1885. Regression towards normality in hereditary stature. J. Anthropol. Inst. 15, 246–263.

Galton, F., 1889. Natural Inheritance. Macmillan, London.

Gauss, C.F., 1866. Theoria interpolationis methodo nova tractata Werke band, Vol. 3. Königliche Gesellschaft der Wissenschaften, Göttingen, pp. 265–327.

Good, B.H., Desai, M.M., 2014. Deleterious passengers in adapting populations. Genetics 198, 1183–1208.

Haldane, J.B.S., 1931. A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. Proc. Camb. Phil. Soc. 27, 131–136.

Henderson, C.R., 1953. Estimation of variance and covariance components. Biometrics 9, 226–252.

Hill, W.G., 1982. Rates of change in quantitative traits from fixation of new mutations. Proc. Natl. Acad. Sci. USA 79, 142–145.

Hill, W.G., 2014. Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. Genetics 196, 1–16.

Hill, W.G., Kirkpatrick, M., 2010. What animal breeding has taught us about evolution. Annu. Rev. Ecol. Evol. Syst. 41, 1–21.

Jenkin, F., 1867. Origin of species. North Br. Rev. 46, 277–318.

Johnson, T., Barton, N.H., 2005. Theoretical models of selection and mutation on quantitative traits. Philos. Trans. R. Soc. London [Biol] 360, 1411–1425.

Kempthorne, O., 1954. The correlation between relatives in a random mating population. Proc. Roy. Soc. London B 143, 102–113.

Kimura, M., 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc. Natl. Acad. Sci. USA 54, 731–736.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press.

Kirkpatrick, M., Johnson, T., Barton, N.H., 2002. General models of multilocus evolution. Genetics 161, 1727D1750.

Kondrashov, A.S., 1984. On the intensity of selection for reproductive isolation at the beginnings of sympatric speciation. Genetika 20, 408–415.

Kruuk, L.E.B., 2004. Estimating genetic parameters in natural populations using the "animal model". Philos. Trans. R. Soc. London [Biol] 359, 873–890.

Lande, R., 1975. The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet. Res. 26, 221–235.

Lange, K., 1978. Central limit theorems for pedigrees. J. Math. Biol. 6, 59–66.

Loewe, L., Charlesworth, B., 2006. Inferring the distribution of mutational effects on fitness in Drosophila. Biol. Lett. 2 (2), 426–430.

Lush, J.L., 1937. Animal Breeding Plans. Ames, Iowa.

Lynch, M., Walsh, B., 1997. Genetics and Analysis of Quantitative Traits. Sinauer Press.

Magnello, M.E., 1998. Karl Pearson's mathematization of inheritance: from ancestral heredity to Mendelian genetics. Ann. Sci. 55, 35–94.

McDonald, D.R., 2005. The local limit theorem: a historical perspective. JIRSS 4 (2), 73–86.

Meuwissen, T., Hayes, B., Goddard, M., 2013. Accelerating improvement of livestock with genomic selection. Annu. Rev. Anim. Biosci. 1 (1), 221–237.

Pearson, K., 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 1–67.

Pearson, K., 1897. Mathematical contributions to the theory of evolution. On the law of ancestral heredity. Proc. Roy. Soc. London 62, 386–412.

Pearson, K., 1904a. Mathematical contributions to the theory of evolution. XII. On the generalised theory of alternative inheritance with special reference to Mendel's Laws. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 203, 53–86.

Pearson, K., 1904b. A Mendelian view of the law of ancestral inheritance. Biometrika 3, 109–112.

Pearson, K., 1909. The theory of ancestral contributions of a Mendelian population mating at random. Proc. Roy. Soc. London 81, 225–229.

Polechova, J., Barton, N.H., 2005. Speciation through competition: a critical review. Evolution 59, 1194–1210.

Provine, W., 1971. The Origins of Theoretical Population Genetics. Univ. of Chicago Press, Chicago.

Rinott, Y., 1994. On normal approximation rates for certain sums of dependent random variables. J. Comput. Appl. Math. 55, 135–143.

Robertson, A., 1960. A theory of limits in artificial selection. Proc. Roy. Soc. London B 153, 234–249.

Robertson, A., 1961. Inbreeding in artificial selection programmes. Genet. Res. 2, 189–194.

Robertson, A., 1966. A mathematical model of the culling process in dairy cattle. Anim. Prod. 8, 95–108.

Santiago, E., 1998. Linkage and the maintenance of variation for quantitative traits by mutation-selection balance: an infinitesimal model. Genet. Res. 161–170.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics., Vol. 84. John Wiley & Sons Inc., New York, pp. 1065–1073.

Slatkin, M., 1970. Selection and polygenic characters. Proc. Natl. Acad. Sci. USA 66, 87–93.

Turelli, M., Barton, N.H., 1994. Statistical analyses of strong selection on polygenic traits: What, me normal? Genetics 1–29.

Weber, K.E., Diggins, L.T., 1990. Increased selection response in larger populations. II. Selection for ethanol vapour resistance in Drosophila melanogaste at two population sizes. Genetics 125, 585–597.

Weissman, D.B., Barton, N.H., 2012. Limits to the rate of adaptation in sexual populations. PLoS Genet. 8 (6), e1002740. http://dx.doi.org/10.1371/journal.pgen.1002740.

Willi, Y., van Buskirk, J., Hoffmann, A.A., 2006. Limits to the adaptive potential of small populations. Ann. Rev. Ecol. Evol. Syst. 37, 433–458.

Wray, N.R., 1990. Accounting for mutation effects in the additive genetic variance-covariance matrix and its inverse. Biometrics 177–186.

Wright, S., 1935. The analysis of variance and the correlation between relatives with respect to deviations from an optimum. J. Genet. 30, 243–256.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. Nature Genet. 42, 565–569.

Yule, G.U., 1902. Mendel's laws and their probable relations to inter-racial heredity. New Phytol. 1, 222–238.

Yule, G.U., 1907. On the theory of inheritance of quantitative compound characters on the basis of Mendel's laws. In: Report of the Conference on Genetics. Royal Horticultural Soc..

Zhang, X.S., Hill, W.G., 2005. Predictions of patterns of response to artificial selection in lines derived from natural populations. Genetics 169, 411–425.