MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2073–2083 doi: 10.1111/mec.13082

INVASION GENETICS: THE BAKER AND STEBBINS LEGACY

Information entropy as a measure of genetic diversity and evolvability in colonization

TROY DAY*†

*Department of Mathematics and Statistics, Jeffery Hall, Queen's University, Kingston, ON K7L 3N6, Canada, †Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada

Abstract

In recent years, several studies have examined the relationship between genetic diversity and establishment success in colonizing species. Many of these studies have shown that genetic diversity enhances establishment success. There are several hypotheses that might explain this pattern, and here I focus on the possibility that greater genetic diversity results in greater evolvability during colonization. Evaluating the importance of this mechanism first requires that we quantify evolvability. Currently, most measures of evolvability have been developed for quantitative traits whereas many studies of colonization success deal with discrete molecular markers or phenotypes. The purpose of this study is to derive a suitable measure of evolvability for such discrete data. I show that under certain assumptions, Shannon's information entropy of the allelic distribution provides a natural measure of evolvability. This helps to alleviate previous concerns about the interpretation of information entropy for genetic data. I also suggest that information entropy provides a natural generalization to previous measures of evolvability for quantitative traits when the trait distributions are not necessarily multivariate normal.

Keywords: adaptation, evolution, evolutionary rescue, heterozygosity, information theory, invasive species, polymorphism

Received 3 October 2014; revision received 8 January 2015; accepted 8 January 2015

Introduction

The publication of the symposium volume *The Genetics of Colonizing Species* by Baker & Stebbins in 1965 was a landmark in the study of species invasions and the colonization of new habitats. A great deal of the published discussion of the symposium and the accompanying publications themselves centred on understanding those characteristics of species that make for good colonizers. For example, rapid growth, plasticity, short generation times and the ability to self have all been suggested as traits that increase colonization ability.

In addition to examining the attributes of individuals that make for good colonizers, one might also consider population-level characteristics that increase the likelihood of colonization success. Several of the contributions in Baker & Stebbins (1965) follow up ideas along this theme as well, but since that time (perhaps because of the influence of that volume), there has been an ever increasing interest in studies of this sort.

One particular area of focus has been the hypothesis that increased trait variation (e.g. genetic diversity) leads to increased colonization success (Hughes *et al.* 2008; Lee & Gelembiuk 2008; Forsman 2014). For example, Forsman (2014) recently surveyed the literature and identified 18 experimental studies on animals and plants that manipulated the genetic (and/or phenotypic) diversity of founder groups and then assessed the effect on establishment success. Of the 18 studies, all but one reported a significant positive relationship between establishment success and genetic diversity. González-Suárez *et al.* (2015) compiled observational data for 511 invasion events involving 97 different mammalian species. Interestingly, they found that establishment success was positively associated with intraspecific variation in adult body mass but negatively associated with variation in neonate body mass.

Correspondence: Troy Day, Fax: 613 533 2964;
E-mail: tday@mast.queensu.ca

Along similar lines, González-Suárez & Revilla (2013) used a large mammalian data set to show that the risk of extinction (for example due to habitat change) tends to decrease as intraspecific variation in adult body mass, litter size and age at sexual maturity each increases.

The above studies suggest that intraspecific diversity in some traits might enhance colonization success or reduce the likelihood of extinction if the environment changes. As has been noted previously, however, there are many possible explanations for this pattern (González-Suárez & Revilla 2013; Forsman 2014; González-Suárez *et al.* 2015). For example, high diversity might result in a high probability that pre-adapted types are already common among the colonizing individuals. Similarly, if some form of niche complimentary plays a role in colonization success, then high diversity might result in a high probability that the appropriate set of types is present among the colonizers (see also Loreau & Hector 2001).

Another explanation is that some form of evolutionary adaptation is required for successful establishment and that high (genetic) diversity results in a greater likelihood of such adaptation as opposed to extinction (Willi *et al.* 2006; Lee & Gelembiuk 2008). Dlugosch & Parker (2008) have shown that, across many species, colonizing populations have reduced genetic diversity when measured using discrete molecular markers. This would therefore lead one to suspect that the evolvability of such populations might often be compromised. When measured for quantitative traits, however, this pattern was less apparent. Furthermore, they suggest that many quantitative traits have evolved rapidly in colonizing populations despite the fact that diversity is sometimes reduced. Thus, the importance of this evolutionary hypothesis remains unclear.

Given the variety of hypotheses for diversity-colonization success relationships, studies of such patterns have unsurprisingly employed a wide variety of measures of diversity. For example, the focus of most of the 18 studies examined by Forsman (2014) was on genetic diversity, and most quantified genetic diversity as the number of genotypes present in the colonizers (e.g. Reusch *et al.* 2005 with *Zostera marina*, Wang *et al.* 2012 with *Spartina alterniflora*, Agashe 2009 with *Tribolium castaneum*, Ellers *et al.* 2011 with *Orchesella cincta*, Hovick *et al.* 2012 and Crawford & Whitney 2010 with *Arabidopsis thaliana*, Robinson *et al.* 2013 with *Daphnia magna* and Drummond & Vellend 2012 with *Taraxacum officinale*). However, several studies used proxies for genetic diversity such as effective population size (Newman & Pilson 1997 with *Clarkia pulchella*), level of inbreeding or relatedness (Leberg 1990 with *Gambusia holbrooki*, and Gamfedlt *et al.* 2005 with *Balanus improvisus*), degree of multiple mating in the parents of founders (Mattila & Seeley 2007 with honey bees) or phenotypic polymorphism (Wennersten *et al.* 2012 with *Tetrix subulata*). Finally, some studies directly measured heterozygosity or levels of polymorphism (Porcaccini & Piazzi 2001 with *Posidonia oceanica*, Markert *et al.* 2010 with *Americamysis bahia* and Martins & Jain 1979 with *Trifolium hirtum*).

The most suitable measure of diversity to use is presumably determined, in part, by the mechanisms that translate diversity into increased colonization success. After all, it will be these mechanisms that determine which aspects of genetic diversity are most important. Therefore, it might prove useful to examine how best to measure diversity in the context of different hypotheses. That is the purpose of this study. I focus on the hypothesis that greater genetic diversity leads to a greater evolvability in the colonizing population, and I ask how should evolvability be measured in the context of colonization studies such as those mentioned above? I will show using some relatively simple evolutionary considerations that Shannon's measure of the information entropy of the standing genetic variation is a natural measure of evolvability.

Quantifying evolvability

Although measures of evolvability has been derived in previous studies (Houle 1992; Hansen 2003; Hansen & Houle 2008), these focus on quantitative traits and restrict attention to cases where the trait follows a normal (or a multivariate normal) distribution. As can be seen from the above list of experimental studies, much of the work on the diversity-colonization success relationship has focused on molecular genetic data or discrete phenotypic traits. It is not clear how previously proposed measures of evolvability might be adapted to this situation, and therefore, I will derive a measure of evolvability by essentially 'starting from scratch'. Interestingly, the measure of evolvability obtained has some features that make it a potential generalization of previous measures of evolvability used for quantitative traits, but for situations where multivariate normality no longer necessarily holds (see Box 1).

I begin by considering the colonization of a new habitat. I use the terms 'native population' and 'native habitat' in reference to those of the source of the colonists. Similarly, I use the term 'novel population' and 'novel habitat' in reference to the colonists. Although this terminology is motivated by the study of colonizing species, the considerations below apply equally to other instances of adaptation to novel environments. For example, this includes instances of lab adaptation as well as adaptation to changes in the environment.

Box 1. The relationship between measures of evolvability

There are two useful ways to compare the measure of evolvability developed here with that proposed for quantitative traits by Houle (1992) and Hansen & Houle (2008).

The most direct comparison can be made by using the general measure of evolvability in eqn (4). This measure makes no assumptions about the strength or form of selection or about the form of the target distribution. Furthermore, in the case of a continuous random variable X representing the breeding value of a quantitative trait, and target and native probability density functions given by $p^*(x)$ and $p(x)$, respectively, the corresponding quantity D can be computed as

$$D(p^*||p) = \int p^*(x) \log_b \left(\frac{p^*(x)}{p(x)} \right) \mathrm{d}x.$$

Therefore, we can compute the evolvability in eqn (4) for any distribution of interest.

In the context of quantitative traits, we restrict attention to those specific types of selection (and mutation) that are compatible with both the native and the target distributions being Gaussian and having a common variance. Under these conditions, the evolutionary change that occurs in going from the native distribution $p(x)$ to the target distribution $p^*(x)$ is completely described by the difference in the means of these two distributions. In this case, $D(p^*||p)$ can be calculated as

$$D(p^*||p) = \frac{1}{2} \frac{(\Delta \bar{x})^2}{v}$$

where $\Delta \bar{x}$ is the difference in the means and v is the (common) variance.

Now if we use the mean-standardized nondimensional breeder's eqn from Hansen & Houle (2008), we have $\Delta \bar{x} = I_A \beta$ where I_A is the mean-standardized variance of the breeding value distribution and β is the mean-standardized selection gradient. Thus, we have

$$D(p^*||p) = I_A \frac{\beta^2}{2}$$

Hansen and Houle proposed $\sqrt{I_A}$, or equivalently I_A, as a measure of evolvability for such traits. The above eqn shows that this is a special case of eqn (4) where the strength of selection is measured by $S(p_i^*, p_i) = \beta^2/2$.

A second type of comparison with previous measures of evolvability can be made by considering how evolvability as measured by information entropy in eqn (7) (which assumes truncation selection and no knowledge of the selective regime in the novel habitat) compares with previous measures if the native distribution is a continuous distribution of breeding values. Although it is well-known that there is no equivalent measure of information entropy for continuous random variables, Shannon himself defined the entropy in such cases by *analogy* with the discrete case. In particular, he used

$$-\int p(x) \log_b p(x) \mathrm{d}x$$

as the information entropy for a continuous random variable X (Shannon & Weaver 1949). The above expression is usually referred to as differential entropy to distinguish it from the right-hand side of eqn (7).

Interestingly, for a mean-standardized normal distribution of breeding values, the differential entropy can be calculated as $\ln(\sqrt{I_A}\sqrt{2\pi e})$ where e is the base of the natural logarithm. Therefore, for normally distributed quantitative traits, information entropy reduces to a simple transformation of the coefficient of variation $\sqrt{I_A}$. Thus, in this special case, we again obtain a measure of evolvability that is effectively the same as that proposed for quantitative traits. This lends support to the idea that information entropy provides a type of generalization of other measures of evolvability for cases where the distribution of breeding values is not Gaussian.

When a species colonizes a new area, often the selective conditions differ from those of the native population. For instance, the new area might have a different temperature profile, salinity or photoperiod, in addition to a different set of biotic factors such as predation, competition and parasitism. Consequently, natural selection will favour different alleles in the novel habitat as compared with the native habitat. For a newly colonizing population to be successful, it must therefore not only overcome the demographic challenges associated with a small pop-

ulation size, but also potentially evolve adaptations to its new habitat if it is to avoid extinction.

To make things concrete, consider a single locus with m potentially segregating alleles (extensions to multiple loci will be considered briefly in the discussion). The probability distribution of alleles in the native population will be referred to as the 'native distribution'. The novel habitat will typically select for a different distribution of alleles, and I refer to this new distribution as the 'target distribution'. I will define the target distribution more precisely later, but for the time being, it should simply be viewed as the allelic distribution that is favoured by selection in the new habitat.

For successful adaption to the new habitat (as opposed to extinction), it might be necessary for the native distribution to evolve into the target distribution. We would therefore like a way of quantifying how easy it is for this to happen. I will restrict attention to the simplest case where the alleles favoured in the novel habitat are present in the standing genetic variation of the native population. This is effectively the same type of assumption that is made in the analyses of evolvability for quantitative traits (Hansen & Houle 2008). Extensions to situations where the favoured alleles must first arise through mutation in the novel habitat are considered briefly in the discussion.

Now consider a native distribution and the target distribution to which it must evolve. How easy is it for this to occur through selection? If the native distribution can achieve target distributions that are very different from itself, then, all else equal, we would say that the native distribution is highly evolvable. At the same time, however, if the native distribution can only achieve such very different target distributions if selection is extremely strong, then we might say that its evolvability is quite low. Thus, I seek a measure that quantifies the magnitude of the change in distribution that can be achieved *per unit strength of selection*, as has been used in other analyses (Hansen & Houle 2008).

The first step is to quantify how much change occurs in going from the native to the target distribution. Suppose p_i is the native distribution of allele frequencies and p_i^* is the target distribution. How can we measure the change achieved when p_i evolves into the distribution p_i^*, from the standpoint of selection and evolution? If the target distribution can be obtained simply by randomly sampling the native distribution, then, in a sense, no change is necessary to go from p_i to p_i^* because the two distributions are effectively the same from the standpoint of selection. This is because no selection (i.e. no biased sampling) is required to obtain the target distribution. On the other hand, if it is very unlikely to obtain the target distribution by randomly sampling the native distribution, then a great deal of change is required to go from p_i to p_i^* because the two distributions would then be very different from the standpoint of selection. Thus, we can use the probability of obtaining the target distribution from the native distribution via random sampling as an evolutionarily relevant measure of the change achieved in evolving from one into the other through selection.

To formalize this idea, we need to calculate the probability of obtaining the target distribution when sampling from the native distribution. I begin by first considering a sample of size n (shortly I will take the limit as $n \to \infty$). To work with a sample of size n, we first need to characterize the target distribution for a sample of size n. Let A_i be random variables denoting the number of alleles of type i that are obtained when drawing a sample of n alleles from the target distribution (with $\sum_i A_i = n$). Now, for any sample obtained from the target distribution, we can calculate the probability of obtaining this same collection of alleles when sampling n alleles from the native distribution. Denoting this probability by Z_n, it is given by the multinomial probability:

$$Z_n = \frac{n!}{\prod_i A_i!} \prod_i p_i^{A_i} \qquad \text{(eqn1)}$$

The probability Z_n in eqn (1) is, itself, a random variable because the A_i are random variables. Different samples from the target distribution will, by chance, result in different numbers of each type of allele, A_i. Consequently, different samples will result in different probabilities Z_n of obtaining the sample when drawing from the native distribution. To obtain a 'typical' value for Z_n, we can calculate an average of Z_n over the draws in the sample for a large sample size n. The multiplicative nature of (1) means that a 'typical' value of Z_n is best characterized by its geometric average, $(Z_n)^{1/n}$. This represents the (geometric) average probability across all draws in a sample, and it takes a value between 0 (the distributions are very different) and 1 (the distributions are effectively the same).

The quantity $(Z_n)^{1/n}$ is also a random variable, but we expect that, from the law of large numbers, it will not vary much if the sample size n is large. Before formally considering this limit, however, I first take the negative logarithm of $(Z_n)^{1/n}$. Denoting the resulting quantity by L_n gives

$$\begin{aligned} L_n &= -\log_b \left(\frac{n!}{\prod_i A_i!} \prod_i p_i^{A_i} \right)^{1/n} \\ &= -\frac{1}{n} \log_b n! + \frac{1}{n} \sum_i \log_b A_i! - \frac{1}{n} \sum_i A_i \log_b p_i \end{aligned} \qquad \text{(eqn2)}$$

where I have left the base of the logarithm unspecified. L_n is a non-negative random variable, and as the native

and target distributions become increasingly different, L_n takes on increasingly large positive values.

The use of the logarithm in (2) has two advantages. First, as will be seen below, the strength of selection is often naturally measured on a additive scale and using the logarithm here means that we are then measuring the change in distribution on an additive scale as well.

Second, using the logarithm provides a convenient way to interpret the change achieved in evolving from the native to the target distribution. The base b is arbitrary and so it can be viewed as setting the scale of measurement. In particular, a value of $L_n = 0$ means that there is a (average) probability of 1 of obtaining the target distribution by randomly sampling the native distribution. A value of $L_n = 1$ means that there is a (average) probability of $1/b$ of obtaining the target distribution. Similarly, a value of $L_n = 2$ means that there is a (average) probability of $1/b^2$ of obtaining the target distribution, and so forth.

Relative values of L_n also have meaningful interpretations. For example, suppose two populations differ by one unit in the amount of change, L_n, that occurs when evolving from the native into the target distribution, and consider working with the base $b = 10$. Then, the population with the larger value of L_n will have evolved ten times as much as the other population in the sense that the probability of it giving rise to the target distribution simply through random sampling will have increased ten times more than that of the other population.

With these preliminaries, we can now take the limit $n \to \infty$. In this limit, the random variable L_n approaches a fixed (i.e. deterministic) value, and defining $D = \lim_{n\to\infty} L_n$, we obtain

$$D(p_i^* || p_i) = \sum_i p_i^* \log_b \left(\frac{p_i^*}{p_i} \right) \qquad \text{(eqn3)}$$

where I have used the notation $D(p_i^* || p_i)$ to indicate that D depends on both the target and native distributions (the notation '||)' in the argument of D is a convention borrowed from information theory). Formally, the above limit holds 'almost surely' and makes use of the strong law of large numbers, Stirling's approximation $\log_b n! = n \log_b n - n + O(\log_b n)$ and the continuous mapping theorem. Analogous calculations have been used in the context of statistical mechanics and Bayesian statistics (for example, see Jaynes 2003). The quantity $D(p_i^* || p_i)$ in (3) that characterizes the amount of change that occurs in evolving from the native into the target distribution is known in information theory as the Kullback-Leibler divergence between the distributions p_i^* and p_i.

To quantify the evolvability of the native population, we also need to know how strong selection must be in order to evolve the target distribution from the native distribution (Hansen & Houle 2008). Many measures of the strength of selection have been proposed but, as far as I am aware, there is no single measure that is suitable under all possible forms of selection. Therefore, although there is a very general way to measure the amount of evolution that occurs through one generation of selection, there does not appear to be a similarly general way to measure the strength of selection required to cause this evolution. Thus, for the moment, I simply denote the strength of selection required to evolve the distribution p_i^* from p_i in a single generation as $S(p_i^*, p_i)$.

With the above two ingredients in hand, we can now define the evolvability $\mathcal{E}$ of the target distribution from the native distribution, as the amount of change in allelic distribution that occurs in going from one to the other, per unit strength of selection. We have

$$\mathcal{E} = \frac{D(p_i^* || p_i)}{S(p_i^*, p_i)}. \qquad \text{(eqn4)}$$

Box 1 discusses the relationship between this general formulation and the specific measures of evolvability for quantitative traits proposed by Hansen & Houle (2008). The next task is to simplify the general measure of evolvability given by eqn (4) for populations that colonize novel habitats.

Information entropy as evolvability when adaptating to novel environments

Equation (4) is a very general measure of evolvability that can be applied to any distribution of alleles at a single locus and to any target distribution (it can also be extended to account for a continuum of alleles; Box 1). It also highlights an important point – the evolvability of a population depends not only on the allelic distribution of that population p_i, but also on the form of selection as quantified by the difference between p_i and p_i^*. A similar dependence occurs with previous measures of evolvability for quantitative traits as well although selection in such cases is necessarily restricted to specific functional forms (e.g. Hansen & Houle 2008; Chevin 2012).

In many situations, we do not know exactly what the form of selection will be, and therefore, we cannot specify a particular target distribution. For example, this is often the case for populations that colonize novel habitats. In such cases, the best we can do is to calculate an expected evolvability over the different forms of selection that might occur (e.g. see Kirkpatrick 2009; Chevin 2012). To do so, we first specify the set of possible alleles. We then specify a class of target distributions for this set of alleles that captures the different forms of selection of interest. Finally, we calculate the expected

evolvability $\mathbb{E}[\mathcal{E}]$ by averaging (4) over these different target distributions, where each target distribution is weighted by its probability of occurrence.

There are two equivalent ways that we can proceed in this direction. The first starts by specifying a class of fitness functions. These can then be used to obtain a suitable measure of the strength of selection for each target, $S(p_i^*, p_i)$. They can also be used to derive the target distributions by determining the distribution of alleles that is produced by each fitness function. From there, one can compute the quantity $D(p_i^*||p_i)$. The second approach starts by specifying a class of target distributions. These will directly determine the quantity $D(p_i^*||p_i)$ for each target. We then derive a fitness function for each target distribution by determining the function required to produce the target through selection. Finally, these fitness functions can be used to obtain a suitable measure of the strength of selection for each target, $S(p_i^*, p_i)$. I follow the second approach below as the general measure of evolvability in eqn (4) has been developed by fixing attention on the native and target distributions themselves.

From the standpoint of evolvability during colonization, some types of target distributions are perhaps of more interest than others. To the best of my knowledge, all previous measures of evolvability for quantitative traits are based on an assumption that, whatever the nature of selection might be, it favours a particular phenotypic or genotypic value. This is embodied by the assumption that selection is directional in these studies (Hansen & Houle 2008). This also seems like a reasonable assumption for adaptation to a novel habitat, and so I employ a form of directional selection as well.

There are many different ways to model directional selection. There is an 'obvious' choice for which of these to use in the context of quantitative traits, however, because the form that is chosen must be compatible with the assumption that all distributions remain multivariate normal. Thus, most such studies assume (sometimes implicitly) that the fitness of different types is specified by a function from the exponential family.

In the context of eqn (4), there is no similar constraint that dictates our choice for the form of fitness because we are allowing for arbitrary target distributions. Thus to model directional selection for a particular allele k, we are free to choose any target distribution subject only to the constraint that the frequency of allele k increases through selection and that of other alleles decreases (i.e. $p_k^* > p_k$ and $p_i^* < p_i$ for all $i \neq k$). This freedom is both a luxury and a curse as we must restrict things more in order to make further progress, but any such restriction is necessarily somewhat arbitrary.

Perhaps the most obvious further restriction for modelling directional selection is to assume truncation selection. This is the most extreme form of directional selection for a particular allele. It is also the simplest form of directional selection in the sense of having the fewest parameters. Furthermore, it is a natural form of selection that one might impose artificially if attempting to quantify the evolvability of a population by selecting it in different directions.

Another convenient feature of truncation selection is that it permits an unambiguous measure of the strength of selection. If we denote the frequency of any favoured type under truncation selection by q, then under random mating the evolutionary change in q as a result of one generation of selection is given by (Wright 1935)

$$\Delta q = \frac{q(1-q)}{2}\frac{\partial \ln \bar{W}}{\partial q} \qquad \text{(eqn5)}$$

where Δq is the change in q and $\bar{W}$ is the population mean fitness. This is an example of Wright's adaptive topography (Wright 1935; Lande 1976; Barton & Turelli 1987), and the magnitude of the quantity $\partial \ln \bar{W}/\partial q$ (which is sometimes referred to as the selection gradient) provides a widely used normalized measure of the strength of selection.

For these reasons, I proceed with an assumption of truncation selection for a particular allele k. In this case, the target distribution is $p_k^* = 1$ and $p_i^* = 0$ for all $i \neq k$, and the associated fitness function that produces this target distribution is $W_k > 0$ and $W_i = 0$ for all $i \neq k$, where W_i is the marginal fitness of allele i. It is worth-noting though that this form of target distribution applies more generally if, instead of quantifying evolvability over a single generation of selection, we were to quantify it over multiple generations. For example, suppose we assume a more general form of directional selection for allele k where we require only that $W_k > W_i$ for all $i \neq k$. This is no longer truncation selection as all alleles might have nonzero fitness. Under these conditions, if no processes other than selection are occurring, then over time the population will evolve to a distribution concentrated entirely at allele k. In other words, the allelic distribution would converge to $p_i = 1$ if $i = k$ and $p_i = 0$ otherwise. Thus, the same target distribution applies to this more general form of directional selection if we quantify evolvability over a large number of generations.

Under truncation selection, we have $\bar{W} = p_k W_k$ and therefore $S(p_i^*, p_i) = |\partial \ln \bar{W}/\partial p_k| = 1/p_k$. Substituting this and $p_k^* = 1$ and $p_i^* = 0$ for all $i \neq k$ into (4) gives

$$\mathcal{E} = -p_k \log_b p_k \qquad \text{(eqn6)}$$

Equation (6) gives the evolvability of the target distribution from the native distribution when there is truncation selection for allele k. As we do not know *a priori*

which allele will have the highest fitness during colonization, the final step is to calculate the expected evolvability $\mathbb{E}[\mathcal{E}]$ over all possible favoured alleles. To do so, we need to specify the probability that allele k will be the favoured allele. If we have no *a priori* knowledge of which allele will be favoured, the only reasonable assumption is that each allele has equal chance of being the favoured allele. Thus, taking the expectation of (6) over a uniform distribution for k gives

$$\mathbb{E}[\mathcal{E}] = -\sum_k p_k \log_{\hat{b}} p_k \qquad \text{(eqn7)}$$

where the normalization constant $1/m$ has been absorbed into the new base of the logarithm, $\hat{b}$. Equation (7) is Shannon's measure of the information entropy of the native allelic distribution (Shannon 1948). Thus, the information entropy of the native allelic distribution provides a natural measure of the evolvability of a population under truncation directional selection when colonizing a novel habitat (Box 2 provides some intuition for thinking about information entropy).

Discussion and concluding remarks

The results derived here suggest that Shannon's information entropy is a sensible measure of genetic diversity in the context of evolvability in novel habitats. Shannon information is frequently used as a measure of species diversity in the ecological literature, and there have also been several instances of its use in population genetics as a measure of genetic diversity (Sherwin *et al.* 2006; Kosman & Leonard 2007; Sherwin 2010). The earliest such instance that I am aware of is Lewontin's (1972) use of information entropy for quantifying patterns of allelic diversity in humans. Nevertheless, this measure has not seen widespread use in population genetics despite the fact that many software packages can routinely compute this quantity.

One possible reason for a general reluctance to use Shannon information in population genetics has to do with its interpretation (Nei 1975; Hennick & Zeven 1991). In ecology, several different measures of diversity are often used and debated, and the discussions sometimes centre around the phenomenological properties of each measure. For example, discussions often consider questions like whether a measure adequately captures species evenness versus richness, or whether it places 'too much' weight on rare species.

In population genetics, the tradition has been to focus more on the mechanistic interpretation of diversity measures. For example, measures like heterozygosity or percentage polymorphic loci have clear population-genetic interpretations. And as Nei (1975) remarked, Shannon information was '...designed to measure the amount of information in information engineering and is not related to any genetic entity. [As such] ...it is not clear what the ... value of this quantity means in terms of genetic materials'.

The derivation presented here partially addresses this issue of interpretation. It shows that information entropy is a natural measure of evolvability during colonization. As an example, suppose we chose the base $\hat{b} = 10$ and we wish to compare the evolvability of two native populations, A and B, and suppose further that population A has value of $\mathbb{E}[\mathcal{E}]$ that is one unit larger than that of population B. Then, under truncation directional selection, population A is 10 times more evolvable than population B in the sense that, per unit strength of selection, population A can evolve a 10-fold greater likelihood of giving rise to the target distribution through simple random sampling.

Interestingly, of the 18 studies on colonization described in the introduction, the earliest of these (Martins & Jain 1979) was also the only study to employ information entropy as a measure of genetic diversity. Coincidentally, Martins & Jain's (1979) study also appears to have been motivated, in large part, by the symposium volume of Baker & Stebbins (1965). They studied rose clover, *Trifolium hirtum*, and examined the effect of the information entropy of the allelic distribution of colonists over two years. In the first year of the study, they found no effect, but in the second year, they found a rather strong positive relationship between entropy and the establishment success of new roadside colonies.

The measure of information entropy in eqn (7) leaves the base of the logarithm, $\hat{b}$, unspecified. One natural choice is $\hat{b} = m$ where m is the number of alleles potentially segregating. Information entropy is always maximized when all alleles are present at equal frequency, and therefore with this choice of $\hat{b}$, the maximum possible information entropy is 1. This choice is analogous to the use of base 2 logarithms in information theory where the random variables of interest are often binary, taking one of two possible values. Thus, while information entropy is measured in *bits* when using base 2, it is measured in so-called *m-ary* units when using base m.

It should also be emphasized that, when using any measure to compare the evolvability of two populations, one should ensure that the *potential* allelic types in each population are the same. Otherwise the meaning of any comparison is unclear. This does not mean that the same alleles need to be segregating in all populations, but only that the alleles segregating in each population are a subset of the same set of m possible alleles. Furthermore, all else equal $\mathbb{E}[\mathcal{E}]$ will tend to increase as the number of alleles included in a sample increases, and therefore, it is also important to control for sampling effort when making comparisons.

The derivation presented here has focused entirely on quantifying evolvability in terms of the allelic distribution of a single locus. Often, however, data are available for multiple loci. In such cases, there are two ways that one might make use of such data. The first is simply to average the value of $\mathbb{E}[\mathcal{E}]$ over all loci. In fact this was exactly the approach taken by Martins & Jain (1979), and it is analogous to the calculation of average heterozygosity across multiple loci Nei (1975).

The second approach is instead to calculate the *joint* information entropy across all loci. To do so, one would use the distribution of all genotypes of interest and calculate the information entropy of this distribution. In other words, each locus would be viewed as a random

Box 2. Gaining an intuition for information entropy

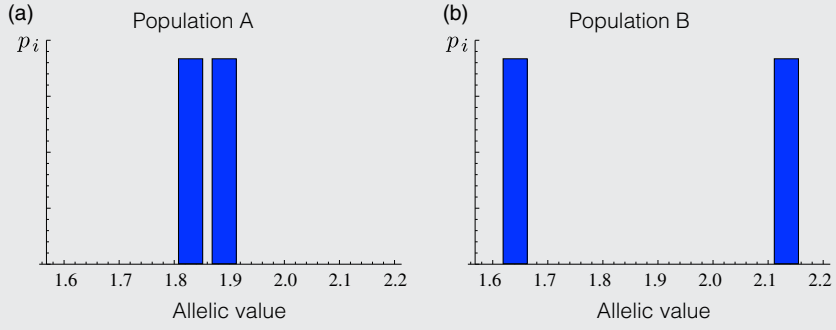
Although information entropy is used frequently in ecology, it is less common in evolutionary biology (although see Frank 2012 for uses of information theory in evolutionary biology that are quite distinct from that explored here). There are two simple ways to begin developing an intuition for what information entropy represents.

The first is rooted in the information-theoretic origins of entropy. Suppose we know the native allelic distribution for some population and imagine that a randomly chosen individual disperses to a novel habitat. For simplicity suppose the organism is haploid. The allelic identity of this single colonist can be viewed as a random variable drawn from the known native allelic distribution. We can then ask the qualitative question, how much new information do we gain about the allelic identity of this colonist if we were to actually measure it?

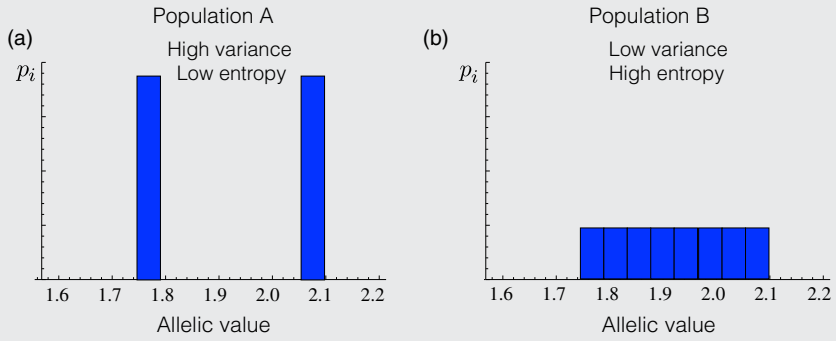
Although the above question is vague in that we have not specified what is meant by 'information', we can make some qualitative progress without being more precise. For example, if the native population contained only a single allelic type, then clearly we would gain no new information by measuring the colonist. This is because we already know with certainty what its identity must be. On the other hand, if there are m potential alleles, and if each of them is equally frequent in the native population, then we would gain a great deal of information by measuring the colonist. This is because its allelic identity prior to measurement is maximally uncertain. And it seems reasonable that we would gain an intermediate degree of new information if the native allelic distribution was somewhere between these two extremes.

Information entropy captures the above qualitative intuition in a precise way. Roughly speaking, it is a measure of the uncertainty of a random variable. Low information entropy corresponds to a low degree of uncertainty in the outcome of a random variable. In such cases, we gain very little new information about a realization of the random variable by seeing its value because there is not much uncertainty in its outcome. On the other hand, high information entropy corresponds to a high degree of uncertainty in the outcome of a random variable. In this case, we gain a great deal of information about a realization of the random variable by seeing its value.

Another useful way to think about the information entropy of a distribution is as a measure of variability. Studies of variance are common in evolutionary biology because, under certain assumptions, genetic variance plays an important role in evolutionary change through natural selection. More generally *genetic variability* is perhaps a more suitable quantity from the standpoint of selection as there needs to be variation for selection to act. Importantly, variability and variance are not always the same thing.



Box Fig. 1 The distributions of allelic values for two populations, A and B. The populations have differing variances and variability as measured by information entropy. Population A has a high variance in allelic value but low variability, while population B has the opposite.



Box Fig. 2 The distributions of allelic values for two populations, A and B. Both populations contain two alleles with equal frequency, and therefore, they have equal variability as measured by information entropy. However, the variance in allelic value of population B is larger than that of population A.

As an example, consider a distribution of discrete allelic values where each allele specifies the value of a quantitative trait like body size. Figure B1 presents distributions from two hypothetical populations that, in an important sense, have the same variability. Both populations have only two alleles and each is equally frequent. As a result, their information entropy is the same. However, the variance in body size in the two populations is very different, with population B having a higher variance. From an evolutionary standpoint, however, there is an important sense in which the two populations are equally evolvable. Therefore variation, as measured by information entropy, can be a more suitable measure of evolvability. Along similar lines, Figure B2 presents distributions from two populations, population A with high variance but low variability (as measured by information entropy) and populations B with low variance but high variability (again measured by information entropy). In this case, it seems reasonable that we would want to classify population B as being more evolvable than population A, again suggesting that variability as measured by information entropy is a more suitable measure of evolvability than variance.

variable, and the genotype distribution would be regarded as the joint distribution of alleles across all loci. The joint information entropy would then be calculated as in eqn (7), but where the summation takes place over all elements of the joint probability distribution. This second approach is preferable in some ways as it gives a total measure of evolvability, accounting for any possible association of alleles through linkage disequilibrium, whereas the first approach gives the average evolvability of a single locus.

As with previous measures of evolvability the measure derived here focuses only on standing genetic variation. While this is likely an important component of adaptation in colonization, novel mutations are likely also important (Schluter & Barrett 2008). Although accounting for this in measures of evolvability is difficult without knowing more about mutational pathways and fitness relationships, we might still make some progress by viewing the target distribution as the distribution obtained from the standing variation that is as close as possible to the distribution favoured by selection. The rationale would be that the closer the population is to the real distribution favoured by selection, perhaps the longer the population can persist before going extinct. As a result, the greater will be the likelihood that the appropriate mutations arise before extinction occurs.

Finally, it is important to emphasize the limitations of information entropy as a measure of evolvability. As eqn (4) shows, the evolvability of any population depends on the form of selection. Consequently, there is no single measure that is appropriate under all conditions. The derivation of information entropy in eqn (7) from eqn (4) rests on two important assumptions: (i) that there is truncation selection in favour of a particular allele and (ii) that all alleles are equally likely to be the favoured allele. If either of these assumptions is relaxed, then a different measure of evolvability might be obtained. For example, if we have reason to believe that certain alleles are more likely to be favoured during colonization than others, then the expected evolvability can be written more generally as $-\mathbb{E}[p_k \log_b p_k]$ where the expectation is taken over an appropriate, nonuniform, distribution. Likewise, relaxing the assumption of truncation selection will typically produce still different measures. Thus, it is important to choose a measure of evolvability that appropriately captures the situation of interest.

Acknowledgements

I thank Mark Blows and David Houle for discussion and David Houle, Steve Frank and two anonymous referees for very helpful feedback on the manuscript. This work was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada. I also thank Wiley publishing for financial support to attend the symposium.

References

Agashe D (2009) The stabilizing effect of intraspecific genetic variation on population dynamics in novel and ancestral habitats. *The American Naturalist*, **174**, 255–267.

Baker HG, Stebbins GL (1965) *The Genetics of Colonizing Species*. Academic Press, Waltham, Massachusetts.

Barton NH, Turelli M (1987) Adaptive landscapes, genetic distance, and the evolution of quantitative characters. *Genetical Research*, **49**, 157–173.

Chevin L-M (2012) Genetic constraints on adaptation to a changing environment. *Evolution*, **67**, 708–721.

Crawford KM, Whitney KD (2010) Population genetic diversity influences colonization success. *Molecular Ecology*, **19**, 1253–1263.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Drummond EBM, Vellend M (2012) Genotypic diversity effects on the performance of *Taraxacum officinale* populations increase with time and environmental favorability. *PLoS ONE*, **7**, e30314.

Ellers J, Rog S, Braam C, Berg MP (2011) Genotypic richness and phenotypic dissimilarity enhance population performance. *Ecology*, **92**, 1605–1615.

Forsman A (2014) Effects of genotypic and phenotypic variation on establishment are important for conservation, invasion, and infection biology. *Proceedings of the National Academy of Science*, **111**, 302–307.

Frank SA (2012) Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *Journal of Evolutionary Biology*, **25**, 2377–2396.

Gamfedlt L, Wallen J, Jonsson PR, Berntsson KM, Havenhand JN (2005) Increasing intraspecific diversity enhances settling success in a marine invertebrate. *Ecology*, **86**, 3219–3224.

González-Suárez M, Revilla E (2013) Variability in life-history and ecological traits is a buffer against extinction in mammals. *Ecology Letters*, **16**, 242–251.

González-Suárez M, Bacher S, Jeschke JM (2015) Intraspecific trait variation is correlated with establishment success of alien mammals. *The American Naturalist*. in press.

Hansen TF (2003) Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *Molecular BioSystems*, **69**, 83–94.

Hansen TF, Houle D (2008) Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, **21**, 1201–1291.

Hennick S, Zeven AC (1991) The interpretation of Nei and Shannon-Weaver within population variation indices. *Euphytica*, **51**, 235–240.

Houle D (1992) Comparing evolvability and variability in quantitative traits. *Genetics*, **130**, 195–204.

Hovick SM, Gumuser ED, Whitney KD (2012) Community dominance patters, not colonizer genetic diversity, drive colonization success in a test using grassland species. *Plant Ecology*, **213**, 1365–1380.

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M (2008) Ecological consequences of genetic diversity. *Ecology Letters*, **11**, 609–623.

Jaynes ET (2003). *Probability theory: the logic of science*. Cambridge University Press, Cambridge, UK.

Kirkpatrick M (2009) Patterns of quantitative genetic variation in multiple dimensions. *Genetica*, **136**, 271–284.

Kosman E, Leonard KJ (2007) Conceptual analysis of methods applied to assessment of diversity within and distance between populations with asexual and mixed mode of reproduction. *New Phytologist*, **174**, 683–696.

Lande R (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*, **30**, 314–334.

Leberg PL (1990) Influence of genetic variability on population growth: implications for conservation. *Journal of Fish Biology*, **37** (supplement A), 193–195.

Lee CE, Gelembiuk GW (2008) Evolutionary origins of invasive populations. *Evolutionary Applications*, **1**, 427–448.

Lewontin R (1972) The apportionment of human diversity. *BMC Evolutionary Biology*, **6**, 381–398.

Loreau M, Hector A (2001) Partitioning selection and complementarity in biodiversity experiments. *Nature*, **412**, 72–76.

Markert JA, Champlin DM, Gutjahr-Gobell R *et al.* (2010) Population genetic diversity and fitness in multiple environments. *BMC Evolutionary Biology*, **10**, 205.

Martins PS, Jain SK (1979) Role of genetic variation in the colonizing ability of Rose Clover (*Trifolium hirtum* All.). *The American Naturalist*, **114**, 591–595.

Mattila HR, Seeley TD (2007) Genetic diversity in honey bee colonies enhances productivity and fitness. *Science*, **317**, 362–364.

Nei M (1975) *Molecular population genetic and evolution*. North-Holland Publishing, Amsterdam.

Newman D, Pilson D (1997) Increased probability of extinction due to decreased genetic effective population size: experimental populations of *Clarkia pulchella*. *Evolution*, **51**, 354–362.

Porcaccini G, Piazzi L (2001) Genetic polymorphism and transplantation success in the mediterranean seagrass *Posidonia oceanica*. *Restoration Ecology*, **9**, 332–338.

Reusch TBH, Ehlers A, Hammerli A, Worm B (2005) Ecosystem recovery after climatic extremes enhanced by genetic diversity. *Proceedings of the National Academy of Science*, **102**, 2826–2831.

Robinson JD, Wares JP, Drake JM (2013) Extinction hazards in experimental *Daphnia magna* populations: effects of genotype diversity and environmental variation. *Ecology and Evolution*, **3**, 233–243.

Schluter D, Barrett R (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

Shannon CE (1948) A mathematical theory of communications. *The Bell System Technical Journal*, **27**, 379–423.

Shannon CE, Weaver WW (1949) *The mathematical theory of communication*. University of Illinois Press, Urbana, IL.

Sherwin WB (2010) Entropy and information approaches to genetic diversity and its expression: genomic geography. *Entropy*, **12**, 1765–1798.

Sherwin WB, Jabot F, Rush R, Rossetto M (2006) Measurement of biological information with applications from genes to landscapes. *Molecular Ecology*, **15**, 2857–2869.

Wang XY, Shen DW, Jiao J (2012) Genotypic diversity enhances invasive ability of *Spartina alterniflora*. *Molecular Ecology*, **21**, 2542–2551.

Wennersten L, Johansson J, Karpestam E, Forsman A (2012) Higher establishment success in more diverse groups of pygmy grasshoppers under seminatural conditions. *Ecology*, **93**, 2519–2525.

Willi Y, Van Buskirk J, Hoffman AA (2006) Limits to the adaptive potential of small populations. *Annual Review of Ecology Evolution and Systematics*, **37**, 433–458.

Wright S (1935) Evolution in populations in approximate equilibrium. *Journal of Genetics*, **30**, 257–266.

T.D. conducted all aspects of the work reported here.
