HAL
open science

On the application of statistical physics to evolutionary biology

N.H. Barton, J.B. Coe

To cite this version:

N.H. Barton, J.B. Coe. On the application of statistical physics to evolutionary biology. Journal of Theoretical Biology, 2009, 259 (2), pp.317. 10.1016/j.jtbi.2009.03.019 . hal-00554594

HAL Id: hal-00554594

https://hal.science/hal-00554594v1

Submitted on 11 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

www.elsevier.com/locate/yjtbi

On the application of statistical physics to evolutionary biology

N.H. Barton, J.B. Coe

PII: S0022-5193(09)00122-2
DOI: doi:10.1016/j.jtbi.2009.03.019
Reference: YJTBI 5498

To appear in: *Journal of Theoretical Biology*

Received date: 7 December 2008
Revised date: 7 March 2009
Accepted date: 10 March 2009

Cite this article as: N.H. Barton and J.B. Coe, On the application of statistical physics to evolutionary biology, *Journal of Theoretical Biology* (2009), doi:10.1016/j.jtbi.2009.03.019

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

On the application of statistical physics to evolutionary biology

N. H. Barton[*,+] and J. B. Coe[*,x]

** Institute of Evolutionary Biology,*
School of Biological Sciences,
University of Edinburgh. Kings Buildings,
Edinburgh EH9 3JT. United Kingdom.

+ Present address:
Institute of Science and Technology,
Am Campus 1,
Klosterneuburg,
Austria A3400.

x Present address:
Institut Curie,
26 Rue d'Ulm,
75248 Paris Cedex 05
France

Accepted manuscript

ACCEPTED MANUSCRIPT

There is a close analogy between statistical thermodynamics and the evolution of allele frequencies under mutation, selection and random drift. Wright's formula for the stationary distribution of allele frequencies is analogous to the Boltzmann distribution in statistical physics. Population size, 2N, plays the role of the inverse temperature, $1/kT$, and determines the magnitude of random fluctuations. Log mean fitness, $\log(\overline{W})$, tends to increase under selection, and is analogous to a (negative) energy; a potential function, U, increases under mutation in a similar way. An entropy, $\boldsymbol{S_H}$, can be defined which measures the deviation from the distribution of allele frequencies expected under random drift alone; the sum $G = E\left[\log(\overline{W}) + U + S_H\right]$ gives a free fitness that increases as the population evolves towards its stationary distribution. Usually, we observe the distribution of a few quantitative traits that depend on the frequencies of very many alleles. The mean and variance of such traits are analogous to observable quantities in statistical thermodynamics. Thus, we can define an entropy, S_Ω, which measures the volume of allele frequency space that is consistent with the observed trait distribution. The stationary distribution of the traits is $\exp\left[2\,N\left(\log(\overline{W}) + U + S_\Omega\right)\right]$; this applies with arbitrary epistasis and dominance. The entropies S_Ω, S_H are distinct, but converge when there are so many alleles that traits fluctuate close to their expectations. Populations tend to evolve towards states that can be realised in many ways (i.e., large S_Ω), which may lead to a substantial drop below the adaptive peak; we illustrate this point with a simple model of genetic redundancy. This analogy with statistical thermodynamics brings together previous ideas in a general framework, and justifies a maximum entropy approximation to the dynamics of quantitative traits.

Keywords: entropy, fitness, random drift, redundancy, information, statistical thermodynamics, Maxwell's Demon

Accepted manuscript

ACCEPTED MANUSCRIPT

There has been a long and varied history of attempts to relate thermodynamics and evolution. From the late nineteenth century up to the present, there has been wide concern that the inevitable increase in entropy apparently contradicts the maintenance of order by living organisms, and its creation by natural selection (Depew & Weber, 1995, Ch. 17). In fact, there is no real difficulty, because organisms are open systems that maintain themselves by exporting entropy to their surroundings (Lotka, 1922; Schrodinger, 1944; Prigogine et al., 1972). A different approach has been to relate thermodynamics to the evolutionary process itself. Most notably, Fisher (1930, p. 36) drew an analogy between the Second Law of thermodynamics, and his "Fundamental Theorem of Natural Selection", which states that the increase in mean fitness caused by selection is proportional to the additive genetic variance in fitness. Fisher's was a purely verbal analogy, with no mathematical relation between the deterministic effects of selection and the statistical process that underpins the Second Law. There have been many subsequent attempts in the same vein, which draw a verbal analogy between the increase in entropy and evolutionary change (e.g. Brooks and Wiley, 1986; Wicken, 1980) or which find quantities analogous to entropy that increase during deterministic evolution (e.g. Ginzburg, 1977; Bomze, 1991; Demetrius, 1997; Baake and Wagner, 2001; Saakian et al., 2006; Wilson, 2008). However, a close analogy can be made when we consider evolution as a stochastic process: classical thermodynamics is based on the aggregate behaviour of a large number of molecules, just as population genetics depends on the aggregate behaviour of many reproducing genes.

Accepted manuscript

ACCEPTED MANUSCRIPT

Recently, Ao (2005, 2008) and Sella and Hirsh (2005a) have drawn attention to this analogy, using a measure of entropy, $\boldsymbol{S_H}$, that was introduced by Iwasa (1988). Our paper sets their work in a wider context, and extends it to cover a broader range of models. The novel feature of our paper is that it extends to the evolution of quantitative traits in genetically variable populations. Quantitative traits that depend on multiple genes define macroscopic states that include large numbers of microstates. This leads to a distinct entropy measure, S_Ω, that gives a general approximation to the dynamics of quantitative traits. We illustrate these ideas with a simple model of genetic redundancy.

Evolution of the allele frequency distribution

Relation with existing work: Sella and Hirsh (2005a) derive a simple expression for the stationary distribution of an asexual population under selection and random genetic drift. They assume that mutations are so rare that the population is almost always fixed for one or other genotype, labelled *i*; for simplicity, they also assume that the rate of mutation from type *i* to type *j* is the same as that in the reverse direction. With these assumptions, their Eq. 7 gives the probability of being fixed for type *i* as:

$$P_i = \frac{W_i^{2N}}{Z} \tag{1}$$

where Z is a normalising constant, W_i is the fitness of type *i*, and there are $N >> 1$ diploid individuals. Moreover, in this stationary state, the flux of substitutions from *i* to *j* is the same as in the opposite direction, a condition known as "detailed balance". (Sella and Hirsh [2005a] derive the stationary distribution for the Moran model and for haploid and diploid versions of the Wright-Fisher model. We focus on the diffusion limit, which approximates a wide range of models when selection and drift are weak).

Sella and Hirsh's (2005a) result had also been derived for a specific model by Berg et al. (2004; see Sella and Hirsh, 2005b), and Aita et al. (2003, 2005) develop a similar approach for understanding evolutionary computation (see Supplementary Information A). The most general analogy with thermodynamics was in fact made much earlier by Iwasa (1988, Eq. 18, discussed below). In fact, this stationary distribution of fixed states (Eq. 1) is a special case of Wright's (1931) stationary distribution of allele frequencies:

$$P[\underline{p}] = \frac{1}{Z}\prod_{k=1}^{n}\left(p_k^{4N\mu_{P,k}-1}\, q_k^{4N\mu_{Q,k}-1}\right)\overline{W}^{2N} \tag{2}$$

Wright's formula gives the stationary distribution of allele frequencies, $\underline{p} = \{p_1, \ldots, p_n\}$, for a sexually reproducing population with *n* loci, each with two alleles at frequency q_k, p_k at locus *k*. The rates of mutation to these alleles are $\mu_{Q,k}$, $\mu_{P,k}$ respectively, and the mean fitness of the population is $\overline{W}$. The alleles may interact in their effects on fitness, but the fitness of each genotype must be constant. The main restriction on Wright's formula is that there must be no statistical associations amongst alleles at different loci ("linkage equilibrium"), so that the state of the population can be described simply by the list of allele frequencies, $\underline{p}$. This will be a good approximation when recombination is faster than selection and drift, as is usually the case for outcrossing sexual populations. If there are multiple alleles at each locus, then the matrix of mutation rates must take a special form that allows a stationary distribution with detailed balance.

Sella and Hirsh's (2005) result (Eq. 1 above) gives a distribution across a space of fixed genotypes, whereas Wright's formula (Eq. 2) gives a distribution across the much larger space of allele frequencies. (Provine (1986, Ch. 9) discusses the relation between these alternative views of the 'fitness landscape'). The former result arises in the limit of low mutation rates ($4N\mu << 1$), when the distribution of allele frequencies is clustered around states of fixation ($p_k = 0$ or $1\ \forall\ k$). Integrating over these regions, we obtain $P_i \sim W_i^{2N} / \prod_k (4N\mu_{i,k})$, where *i* labels the 2^n genotypes, and $\mu_{i,k}$ is the mutation rate away from the allele

Accepted manuscript

ACCEPTED MANUSCRIPT

that is fixed at locus k. This is a generalization of Eq. 1 to asymmetric mutation $(\mu_P \neq \mu_Q)$. This result can be derived more directly by considering the probability of fixation of rare mutations (e.g. Iwasa, 1988, p. 271). Just as mutation reduces fitness below the "adaptive peak" in deterministic models of "quasispecies" (Eigen, 1971), so random drift reduces mean fitness below the maximum possible. For specific models, this "drift load" is proportional to $\frac{1}{2N}$ (Kimura and Ohta, 1970; Sella and Hirsh, 2005a), but this is not a general result (Supplementary Information B).

Sella and Hirsh's model of an asexual population is equivalent to one of multiple alleles at a single genetic locus. However, since they consider substitutions that occur one at a time, recombination is irrelevant, and the results also apply with free recombination in this low-mutation rate limit. For simplicity, we deal with two alleles at each of multiple loci. This generalises to multiple alleles at each locus, provide that the mutation rates take a special form that leads to a stationary state with detailed balance. Sella and Hirsh's model of mutation satisfies this constraint.

We have contrasted Sella and Hirsh's model, in which populations jump between fixed states, with the more general case of polymorphic populations that are described by their allele frequencies. There is a still more general case, in which populations are described by their genotype frequencies - that is, by their allele frequencies plus all the linkage disequilibria that describe associations between alleles at different loci. With n biallelic loci, the distribution across fixed genotypes is described by a vector of 2^n numbers. With n allele frequencies, we must follow the distribution across an n-dimensional space; and with 2^n genotype frequencies, we must follow the distribution across a $2^n - 1$ dimensional space. This latter case is difficult, because in general, there is no detailed balance in the stationary state, and it is not possible to write down the stationary density explicitly. However, Ao (2005, 2008) has made some progress on this problem (see Discussion).

Entropy and free fitness: Under quite general conditions, a measure can be defined which increases as a Markov process approaches its stationary distribution, P_0:

$$\triangle E\left[\log\left[\frac{P_0}{P}\right]\right] \geq 0 \tag{3}$$

Here, $E[]$ represents the expectation over a distribution of states, $P[\underline{p}]$, which changes through time. This quantity is (minus) that defined by Boltzmann in his H-theorem (Boltzmann, 1872; Keizer, 1987, p. 76; Le Bellac et al., 2004, p. 59); it is also known as the relative entropy, or the Kullback-Leibler distance between P and P_0 (Kullback, 1987). One way to understand it is in terms of statistical inference (Jaynes, 1983): it is equal to the expected log likelihood of the hypothesis that the population is sampled from P_0, relative to the likelihood that it is drawn from the actual distribution P. When P differs from P_0, this is negative, since it is on average less likely that the population is drawn from P_0 than that it is drawn from the actual distribution, P. It reaches a maximum at zero when $P = P_0$.

In an evolutionary context, Iwasa (1988, Eq. 7) defined a *free fitness, G*, by taking the expectation in Eq. 3, divided by $2N$. Substituting for the stationary distribution in Eq. 2, and dropping the constant Z gives:

$$\begin{aligned} G &= \frac{1}{2N} E\left[\log\left[\frac{1}{P}\prod_{k=1}^{n}\left(p_k^{4N\mu_{P,k}-1}\, q_k^{4N\mu_{Q,k}-1}\right)\overline{W}^{2N}\right]\right] = \\ & E\left[\log[\overline{W}]\right] + E[U] + \frac{1}{2N} S_H \end{aligned} \tag{4}$$

where $U = 2\sum_{k=1}^{n}\left(\mu_{P,k}\,\mathrm{Log}[p_k] + \mu_{Q,k}\,\mathrm{Log}[q_k]\right)$ is a measure of genetic diversity, and $S_H = -E\left[\mathrm{Log}\left[P\left(\prod_{k=1}^{n} p_k\, q_k\right)\right]\right]$ is a measure of entropy. (Sella and Hirsh (2005a, Eq. 14) and Iwasa (1988,

Accepted manuscript

ACCEPTED MANUSCRIPT

Eq. 19) give a form of this free fitness for the special case where populations are close to fixation, by dropping the second term, E[U]). We label the entropy measure by H because it is close to the quantity defined by Boltzmann's (1872) H-theorem; as we discuss below, it is also proportional to Shannon's (1948) *information entropy*. When averaged over an ensemble of independently evolving populations, the free fitness gives a measure of the deviation from the stationary distribution, whose expectation can never decrease. The three terms represent the effects of selection, mutation, and drift, respectively. Selection tends to increase mean fitness; mutation tends to push allele frequencies towards an equilibrium value at which the second term, $E[U]$, is maximised; and random drift tends to spread out the distribution, thus increasing the last term. This last measure of the dispersion of the probability distribution is analogous to an entropy, and the factor $\frac{1}{2N}$ to a temperature.

We have written the free fitness in a slightly different form from Iwasa (1988), who included the factor $E[\mathrm{Log}[\prod_{k=1}^{n} p_k q_k]]$ with the second term rather than the last. We have a free choice as to how to partition across the three terms: our choice was motivated by requiring that the terms correspond to the effects of selection, mutation, and drift, respectively. We discuss an alternative choice in Supplementary Information C.

Application to quantitative genetics

Microstates and macrostates: Quantitative genetics describes the evolution of complex traits that depend on multiple interacting genes. Typically, we observe the distribution in the population of a trait z, but do not know the allele frequencies that influence it. Thus, the central problem of quantitative genetics is closely analogous to statistical mechanics, which follows macroscopic variables such as energy and pressure, without detailing the dynamics of individual molecules. We believe that the analogy between statistical physics and population genetics is most fruitful in this context.

Entropy was originally developed in classical thermodynamics, in terms of macroscopic flows of heat and work. Since population genetics contains no conserved quantity analogous to energy, this classical concept cannot be applied. The measure defined by Sella and Hirsh (2005a, Eq. 14) and Iwasa (1988, Eq. 7) is a measure of statistical entropy that was introduced by Boltzmann (1872) in his H-theorem; it is proportional to Shannon's (1948) information entropy. This measure, which we denote S_H, describes the dispersion of microstates, and does not involve any macroscopic variables. A third concept, which we denote S_Ω, links macroscopic with microscopic descriptions; it was defined by Boltzmann (1877), and is proportional to the logarithm of the number of microstates consistent with a given macroscopic state. (This measure of entropy was introduced in population genetics by Barton (1989) and Barton and Rouhani (1993)). Other things being equal, systems will tend to evolve towards macroscopic states with higher S_Ω simply because these are consistent with more microstates. This measure of entropy is often presented in terms of a count of discrete states. However, in both population genetics and classical physics, we deal with a continuous space of allele frequencies or of positions and momenta. In this continuous setting, entropy is defined by the volume of state space consistent with an infinitesimal volume of macroscopic states; it requires that we define a metric on both macrostates and microstates. S_Ω and S_H are distinct: S_Ω is a function of the values of the macroscopic variables, whereas S_H is a functional of the distribution of microstates. However, the distribution of microstates that maximises S_H, for given macroscopic values, converges to S_Ω when traits fluctuate close to their expectations (Supplementary Information E).

Suppose that a polygenic trait is approximately normally distributed. Then, the state of the population is described by the trait's mean and variance, $\{\bar{z}, v\}$. If selection acts solely on the trait, then mean fitness is a function $\overline{W}[\bar{z}, v]$. The effect of mutation is described by a third variable, $U = 2\mu \sum_{k=1}^{n} \log(p_k q_k)$, which would not usually be observable. (We assume two alleles per locus, and for simplicity, symmetric mutation,

Accepted manuscript

ACCEPTED MANUSCRIPT

$\mu = \mu_P = \mu_Q$). The stationary distribution of $\{\bar{z}, v, U\}$ is then obtained by integrating over the distribution of allele frequencies (Eq. 2), conditional on these values:

$$P[\bar{z}, v, U] = \frac{1}{Z} \overline{W}[\bar{z}, v]^{2N} e^{2NU} e^{S_\Omega} \tag{5}$$

where $S_\Omega[\bar{z}, v, U] = \mathrm{Log}\left[\int_{\bar{z},v,U} \frac{d\underline{p}}{\prod_{k=1}^{n} p_k q_k}\right]$ involves an integral over the space of allele frequencies, conditioned on values of $\bar{z}$, v, U. Note that although the unconstrained integral diverges as $\int p^{-1} d\underline{p}$ (corresponding to accumulation at fixed states, $p_k = 0$ or 1), it is well-defined when the constraint on U requires that there be variation at every locus. The normalising constant Z is analogous to a partition function, and can be used to recover information about the expected state of the system. For example, if $\log[\overline{W}]$ is proportional to a selection coefficient s, then $\partial \log[Z]/\partial \log[s] = E\left[\log[\overline{W}]\right]$, and if U is proportional to a mutation rate μ, then $\partial \log[Z]/\partial \log[\mu] = E[U]$ (Supplementary Information D).

Although we assume normality, so that mean fitness is a function solely of $\bar{z}$, v, we have not assumed that the trait, z, is given by the sum of effects of the different genes (i.e., that it is additive). Normality would follow from the assumptions of additivity and linkage equilibrium, but even if genes do interact to determine the trait, the trait may still follow an approximately Gaussian distribution. For example, the trait might be the sum of effects of interacting pairs of genes, or more generally, interactions might fluctuate in sign so that the overall distribution remains Gaussian (Turelli and Barton, 2006). Equation 5 applies with arbitrary patterns of interaction; however, complex interactions would make calculation of S_Ω intractable.

Directional selection on an additive trait: The stationary distribution of the macroscopic variables (Eq. 5) expresses the tension between the three different evolutionary forces of selection, mutation and drift (Iwasa, 1988; Barton, 1989; Barton and Rouhani, 1993; Sella and Hirsh, 2005a). To illustrate this point, consider the simple model of directional selection on an additive trait. Suppose that individual fitness depends on $z = \sum_{k=1}^{n} \gamma_k \left(X_k + X_k^* - 1\right)$, where γ_k is the effect of the k'th locus, and X_k, $X_k^* = 0, 1$ are the states of the two copies of the k'th locus in diploids. Individual fitness is $e^{\beta z}$, corresponding to multiplicative selection $s_k = \beta \gamma_k$ at the k'th locus. Assuming normality, we have $\log[\overline{W}] = \beta \bar{z} + \frac{\beta^2 v}{2}$. When selection is weak (β small), as we assume, the second term $\frac{\beta^2 v}{2}$ is negligible.

For simplicity, we assume equal allelic effects ($\gamma_k = 1 \; \forall \; k$), and work with $s = \beta$. The population mean must lie in the range $-n \le \bar{z} \le n$, and the trait variance must be less than $\frac{n}{2}\left(1 - \left(\frac{\bar{z}}{n}\right)^2\right)$; thus, the population lies within the space shown in Figure 1a. With this simple form of directional selection, $\log(\overline{W})$ increases linearly towards fixation of the fittest genotype ($\bar{z} = n = 100$, $v = 0$, at right). Mutation, acting via the second term in Eq. 5, e^{2NU}, makes no direct contribution to the stationary density of z, but the third term, analogous to an entropy, does. In the absence of mutation or selection, the population is equally likely to fix any of the 2^n available genotypes, and so the trait mean follows a binomial distribution tightly clustered around $\bar{z} = 0$ - the state that can be realised in the largest number of ways (dots on horizontal axis in Fig. 1a). When mutation and random drift act in the absence of selection, the stationary density is given by $e^{2NU + S_\Omega}$, and clusters around zero mean and a variance maintained by a mutation-drift balance (central contours in Fig. 1a). Including selection (i.e., multiplying by $\overline{W}^{2N}$) shifts this distribution to the right, towards the maximally fit state (arrow in Fig. 1a).

Fig. 1

In the example of Fig. 1, $2N\mu = 0.1$, and so allele frequencies at individual loci tend to be close to fixation (Fig. 1b). However, with a large number of loci ($n = 100$), the trait mean and variance follow a Gaussian

Accepted manuscript

ACCEPTED MANUSCRIPT

distribution that clusters around their most likely state (Fig. 1a). With a smaller number of loci ($n = 10$, say), the Gaussian approximation is still quite accurate, except where the distribution is close to the edge of its allowable range. In such cases, the exact distribution can be calculated from Eq. 5, as the convolution of n allele frequency distributions. Barton (1989) and Coyne et al. (1997, Appendix) explore such calculations, for a model of stabilising selection on an additive trait.

Figure 2 shows how mean fitness, due to a single locus under directional selection s, is reduced below the maximum possible by mutation and drift. Each panel shows the average allele frequency, $E[p]$, which is directly related to log mean fitness $\left(\log\left[\overline{W}/\overline{W}_{\max}\right] = -2\,ns(1 - E[p])\right)$. Figure 2a shows how the average frequency of the fitter allele increases with population size, for a fixed rate of mutation relative to selection, whilst Figure 2b shows how mean allele frequency depends on $\frac{\mu}{s}$, for given $2\,Ns$. For $2\,Ns >> 1$, $E[p]$ approaches the deterministic limit $p = \frac{1}{2} + \frac{1}{2}\sqrt{1 + \left(\frac{2\,\mu}{s}\right)^2} - \frac{\mu}{s}$ which is approximately $1 - \frac{\mu}{s}$ for $\mu << s$ (asymptotes at right of Fig. 2a; upper curve in Fig. 2b).. In small populations ($2\,Ns << 1$), allele frequency and mean fitness increase linearly with population size, as $E[p] = \frac{2\,Ns}{1+8\,N\mu}$ (linear region at left of Fig. 2a; lower curve in Fig. 2b). When mutation rates are very low ($\mu << s$), populations are almost always close to fixation, and we have $E[p] = 1\big/\left(1 + e^{-4\,Ns}\right)$. In this limiting case, the genetic load is entirely due to random drift, rather than to mutation. In general, however, the effects of drift and mutation on mean fitness cannot be cleanly separated.

Fig. 2

Modifiers of redundancy: The analogy with thermodynamics highlights the point that, other things being equal, populations will drift into mqcroscopic states that can be realised across a larger volume of allele frequency space - that is, which have higher "entropy", S_Ω. This effect can be seen in Figure 1a, where the trait mean will, under random drift alone, tend towards intermediate values, because those can be realised by a larger number of genotypes (dots around $\bar{z} = 0$, $v = 0$). When mutation rates are low ($N\mu << 1$), the expected trait mean is determined by a balance between this effect, and the tendency of selection to maximise fitness. Iwasa (1988, Eq. 20) shows how, if the number of available states increases sufficiently rapidly as fitness decreases, the entropy effect can overwhelm selection.

Accepted manuscript

ACCEPTED MANUSCRIPT

We can illustrate this point with a simple thought experiment, which shows that a modifier allele can invade if it gives high fitness across a wide range of genetic backgrounds, even if it is not present in the fittest possible genotype. Assume that the quantitative trait, z, just depends on the number of '1' alleles in the diploid genotype; with n loci of equal effect, z is defined to range from $-n$ to $+n$.. Suppose that when the '0' allele at the modifier locus is present, fitness depends on n loci as before ($W = e^{sz}$), but when the '1' allele is present, fitness is $e^{sn-\delta}$, independent of z; this robustness comes at a cost to fitness of $\delta > 0$. If both mutation and random drift are weak relative to selection, then the population will be close to fixation for the fittest genotype ($z = n$), and the '1' modifier allele cannot invade. However, if mutation and drift cause the population to spread over a wide range of less fit genotypes, then this allele may have higher fitness than its alternative, when averaged over these states. The distribution of frequencies of the modifier allele is the same as for a single-locus system, with allelic fitnesses $\overline{W}_0 : \overline{W}_1$; this follows directly from our assumption of linkage equilibrium. These marginal fitnesses are simply the mean fitnesses of populations fixed for the alternative alleles at the modifier locus; in this model, $\overline{W}_0 = \exp[ns(2\,E[p] - 1)]$, where $E[p]$ is shown in Fig. 2, and $\overline{W}_1 = e^{sn-\delta}$. The modifier allele is favoured both by mutation and by drift; in the limit of low mutation, where it gains its advantage solely through the 'entropy' term, S_Ω, the condition for invasion is that $e^{sn-\delta} > \exp\left(sn\left(2\,\frac{1}{1+e^{-4Ns}} - 1\right)\right)$, or $\delta < \frac{2ns}{1+e^{4Ns}}$. With large numbers of loci ($n >> 1$), and drift of the same order as, or stronger than, selection ($Ns \lesssim 1$), the entropic effect dominates.

More generally, we can ask what relation between genotype and fitness is likely to evolve. For simplicity, suppose that fitness is some function of an additive trait, z; for example, Fig. 3a compares flat, linear and quadratic relations between log fitness and the trait. We can imagine a locus that carries modifier alleles, with each allele giving a particular relation $W(z)$. With low mutation rates and large population size ($N\mu << 1$, $Ns >> 1$), the single fittest genotype will evolve (upper right of Figs. 3a, b). However, when variation within and between populations is introduced by mutation and drift, then the mean fitness of a modifier allele depends on the fitness averaged across all background genotypes. Specifically, it is just the mean fitness, taken across the stationary distribution (Eq. 2). Moreover, the stationary distribution of the modifier allele frequency is determined by the mean fitness of the modifier alleles, and is given by the single-locus version of Eq. 2. In the example of Fig. 3b, when Ns is small, a flat relation between trait and fitness, and negative (synergistic) epistasis, are favoured.

The general point that the success of an allele depends on its marginal fitness, averaged across the genetic backgrounds on which it finds itself, is well understood: for example, it was emphasised by Dobzhansky (Lewontin, 1981). It has recently been discussed primarily in relation to robustness against mutation (e.g. Schuster and Swetina, 1988, Burch and Chao, 2000, De Visser et al., 2003), and the tendency of mutation to drive populations towards flatter parts of the adaptive landscape has been termed "survival of the flattest" (Wilke et al., 2001; Wilke, 2005). Here, we are simply pointing out that even when mutation is rare, the genetic background will vary if selection is weak relative to random drift, and alleles will be selected for their effects across a range of more or less degraded backgrounds. The effects of random drift have also been studied by Krakauer and Plotkin (2002), who analyse small perturbations from the deterministic equilibrium, and by van Nimwegen et al. (1999), who assume, like Sella and Hirsh (2005), that populations are usually fixed for a single genotype. In a more biological context, Kondrashov (1995) and Lynch (2007) have emphasised the possible consequences of small Ns for genome evolution.

It is not clear that it is sensible to think of "modifiers of epistasis". Unlike classical modifiers of dispersal, mutation or recombination, modifiers of epistasis necessarily have direct effects on fitness, and it is arbitrary which of an interacting set of loci we label as the modifier (Hansen, 2006, p. 138). Regardless of the labelling, at linkage equilibrium, what will evolve will be the set of genotypes that has highest mean fitness, averaged over the distribution generated by mutation and drift. The real difficulty in understanding this evolution of "genetic architecture" is to know what constrains the relation between genotype and fitness (Hansen, 2006). For example, Desai et al. (2007) argue that in the presence of deleterious mutations,

Accepted manuscript

ACCEPTED MANUSCRIPT

selection favours antagonistic (positive) epistasis. However, this result arises because in their model, what is being selected is primarily a reduced deleterious effect of mutations: whether this is accompanied by positive or negative epistasis depends on the constraints that are assumed (Desai et al., 2007, pp. 1008-1009).

In the limit of low mutation rates, the population will almost always be fixed for one or other genotype, and the outcome will be the same, regardless of whether or not there is recombination. We focus on this limit, because we are concentrating on how populations adapt when fitness is lost because of drift, rather than because of mutation. Imagine that the population is initially fixed for an allele that makes fitness insensitive to z, but at some fitness cost (flat line in Fig. 3a). Occasional mutations will cause it to hop between alternative states, following a neutral distribution in which each genotype is equally likely; the trait z will therefore follow a binomial distribution centred on $z = 0$. Now, imagine that an alternative allele is introduced, which increases the fitness of alleles that have high values of z, but reduces the fitness of alleles with intermediate z (straight line in Fig. 3b). This will only be likely to invade when it arises within a population that happens to have sufficiently high z - a very rare event for large numbers of loci, n. However, once the modifier has fixed, the population is likely to evolve towards high z, and will have higher fitness than before, because mutations that reduce z are now strongly deleterious. We can see from this example, that although the stationary distribution tells us that the long-term outcome is that the fitness profile $W(z)$ with highest mean fitness is most likely, it may take a prohibitively long time for this to evolve. (This is similar to the situation with mutation/selection balance and asexual reproduction, where advantageous alleles can only fix if they arise in a sufficiently fit genetic background; Fisher, 1930, p. 122; Charlesworth, 1994; Johnson and Barton, 2002).

Discussion

Analogy with thermodynamics: When a population can be described in terms of allele frequencies, its stationary state under selection, mutation and random genetic drift is given by Wright's (1931) formula (Eq. 2). This immediately suggests a precise analogy with the Boltzmann distribution, with log mean fitness analogous to a (negative) energy, and the inverse of population size analogous to temperature: fluctuations in mean fitness are stronger in a smaller population, in just the same way that fluctuations in energy are stronger at higher temperatures. We can think of selection as causing a systematic increase of those particular alleles that raise fitness $\left(E\left[\Delta \mathrm{p}_i\right] = (p_i q_i/2)\left(\partial \log(\overline{W}) \big/ \partial p_i\right)\right)$. This is analogous to the organised molecular movements required to produce work in classical thermodynamics. In contrast, random drift causes random fluctuations in allele frequencies, analogous to the disorganised motions of heat.

Ao (2005, 2008) has set out this analogy in detail, identifying gradual parameter changes that preserve the stationary distribution as being "reversible" in the thermodynamic sense. It is then possible to make a precise analogy with the Carnot cycle: isothermal changes, in which parameters such as the strength of selection change at constant population size, alternate with adiabatic changes, in which population size and intensive parameters change together (i.e., keeping Ns, $N\mu$ etc. constant), so as to keep the allele frequency distribution the same. However, there is no equivalent to the conservation of energy in population genetics, and it makes no sense to think of exchange of energy with the outside world. Therefore, this line of thinking does not lead to any constraint on the increase in mean fitness that would correspond to the constraint identified by Carnot in classical thermodynamics. Although this analogy is intriguing, it seems to have no biological significance.

If, as is often the case, we observe quantitative traits that depend on genotype in a complex way, then the distribution of these traits is found by averaging over Wright's distribution (Eq.5; Barton, 1989). This gives a product of three terms that correspond to the effects of selection that increases mean fitness; mutation that increases genetic diversity; and an entropy-like effect, that favors those macroscopic states that can be realised through the largest set of allele frequencies. Again, these terms are analogous to those that arise in statistical mechanics, where we see a macroscopic system as an average over very many microscopic states.

Accepted manuscript

ACCEPTED MANUSCRIPT

We also find that the stationary distribution maximises statistical entropy, S_H, given constraints on the expected values of observables (Supplementary Information D),

The analogy with thermodynamics is limited: mean fitness is not conserved in the same way as energy; there is no constraint on the increase in mean fitness, analogous to the Second Law; populations do not tend towards the same size when coupled together, as physical systems would tend towards the same temperature; and there is no principle of equipartition, which spreads populations out evenly over the space of allele frequencies. Rather, the analogy follows from the general properties of Markov processes, which justifies a fundamentally statistical view of entropy (Jaynes, 1983; le Bellac et al., 2004).

The analogy is greatly simplified by the fact that the response of a population to selection is equal to the gradient in log mean fitness, multiplied by a genetic variance which also gives the rate of random fluctuations. We have developed the analogy by assuming linkage equilibrium (so that the population can be represented by its allele frequencies) and by assuming special forms of mutation and selection (so that Wright's formula applies, and the population tends towards a stationary state with detailed balance). When linkage disequilibria are significant, when selection is frequency-dependent, or with more general forms of mutation, the stationary distribution cannot be written down explicitly, and does not cleanly separate into separate factors. Nevertheless, we can still regard the distribution of observed variables, such as the mean and variance of a quantitative trait, as an average over the space of genotype frequencies, and we still expect the population to tend towards states that can be realised in the largest number of ways. Ao (2005, 2008) develops methods for analysing systems that do not tend towards a stationary state, with detailed balance.

The size of the system: In our analogy, the number of genetic loci corresponds to the size of a thermodynamic system - the number of gas molecules, for example. Thus, parameters such as mutation rate, μ, selection gradient, β, and population size, N, that act at each locus can be seen as intensive variables, independent of the size of the system, and analogous to pressure or temperature. Similarly, variables that depend on all the loci (for example, the measure of diversity, U, or a quantitative trait, z) can be seen as extensive variables. With additive traits, this analogy is straightforward: we can define an aggregate trait which is the sum of two traits that depend on different sets of loci, in much the same way that we can make a physical system by combining two smaller systems, and adding up the extensive variables. However, in general we do not assume that quantitative traits are additive; with epistasis, it is harder to make the analogy with physical systems, since there is no quantity such as mass or energy that is conserved, and no simple operation that corresponds to aggregating two systems.

Genetic systems typically are much smaller than familiar physical systems: quantitative traits might depend on hundreds of genes at most, whereas the number of molecules in a macroscopic systems is of order Avogadro's number, $\sim 6\times10^{23}$ mole^{-1}. Thus, we cannot be as confident in describing trait variation by averaging over the underlying genetics as we can be in using classical thermodynamics to describe macroscopic physical systems. Nevertheless, the distribution of quantitative traits clusters quite closely around its expectation even with a modest number of loci (Fig. 1a). Moreover, although we have only a rough idea of how many loci typically contribute to trait variation, it may well be large enough for populations to approach the thermodynamic limit. It is true that in crosses between inbred lines, quantitative trait loci (QTL) with major effects are often detected, accounting for a substantial fraction of genetic variance. However, populations typically respond in a steady and replicable way for as long as selection is applied (Barton and Keightley, 2002), a pattern most naturally explained by the infinitesimal model (Bulmer, 1980), which assumes an indefinitely large number of genes. The success of quantitative genetics in smoothing over the underlying genetics suggests that the analogy with thermodynamics may be fruitful, despite the very different sizes of the systems involved.

Quasispecies: Eigen (1971) modelled a population of sequences evolving under mutation and selection. This model of a "quasispecies" (Eigen and Schuster, 1977) describes the deterministic evolution of an asexual population; it is the focus of an extensive literature, mainly published in physics journals. There is

Accepted manuscript

ACCEPTED MANUSCRIPT

also intense interest within population genetics in the interaction between mutation and selection, but the concerns are broader, including sexual reproduction, recombination, and random drift; Wilke (2005) reviews the relation between these literatures, which have remained largely separate. The quasispecies model is essentially linear, and so its solution can be written explicitly. This leads to a precise analogy with quantum mechanics, in which the inverse temperature corresponds to time: deterministic evolution towards the equilibrium corresponds to cooling towards the ground state (Baake et al., 1997; Baake and Wagner, 2001; Saakian and Hu, 2004). This is quite different from the analogy described here, in which the model is stochastic, with temperature corresponding to the rate of random genetic drift.

Robustness: We have shown that a modifier allele can increase in frequency if it causes high fitness across a wide range of genetic backgrounds, even if it reduces the fitness of the optimal genotype; this effect can be strong, since it increases in proportion to the number of genes involved. This can be thought of as a form of selection for redundancy: the 'entropy' S_Ω which we have defined drives populations towards states that can be realised in many ways, even at the expense of a reduction in maximum fitness (Iwasa, 1988). There has been considerable recent discussion of the evolution of redundancy (e.g. de Visser et al., 2003; Hansen, 2006; Wagner, 2005), stimulated by the surprising discovery that most genes in eukaryotes can be deleted without obvious ill-effects. This issue is closely related to the robustness of organisms to genetic and environmental perturbations, which facilitates evolution of novel features without disruption of existing function. This discussion has focussed on the effects of deleterious mutation; the analogy with thermodynam ics emphasises that random drift also causes populations to spread over sub-optimal states, and so they will tend to adapt to the inevitable presence of drift load as well as mutation load. Indeed, even in the limit where mutation is very rare, modifiers that increase redundancy can still evolve, albeit slowly. Balancing selection also contributes to the maintenance of the ubiquitous genetic diversity that we observe: organisms must evolve to cope with all these sources of variation in the genetic background.

Approximating the dynamics: Although the analogy with thermodynamics gives an intriguing interpretation of Wright's (1931) formula for the stationary distribution, it does not lead directly to new results: the argument drawn from physics that systems will tend towards states with higher entropy corresponds to the straightforward biological argument that some states will be more likely to evolve because they can be generated by a larger set of allele frequency combinations. The thermodynamic interpretation is, however, valuable when extended to approximate the dynamics of quantitative traits. The rate of change of trait means caused by selection and drift is proportional to the additive genetic variance, but the variance itself evolves, in a way that depends on the underlying allele frequencies. In a series of papers, Prugel-Bennett, Rattray and Shapiro approximate these allele frequencies by supposing that they maximise an entropy measure, conditional on the observed distribution of the trait (Shapiro et al., 1994, Rogers and Prugel-Bennett, 2000, Rattray and Shapiro, 2001). Their simulations suggest that this can be a remarkably accurate approach. However, their entropy measure differs from that used here, and is not justified by any evolutionary model; their procedure can be seen as an ad hoc technique for moment closure. In contrast, the entropy measure that we define here is justified in that the stationary distribution is recovered correctly (Supplementary Information D). Provided that the quantitative trait is perturbed from its steady state slowly enough, then the underlying allele frequencies should stay close to their stationary distribution, and the approximation will remain accurate. (Such changes are termed "reversible", by analogy with physical perturbations that are slow enough for the Boltzmann distribution to be maintained; Ao, 2005). Without examination of specific models, however, it is not clear whether this would be a good approximation to more rapidly evolving populations, where the quantitative trait and the allele frequencies evolve on the same timescale. Barton and de Vladar (2009) show that under directional selection, this approach is accurate even in the worst case of an abrupt change; they also outline the application to stabilising selection on a quantitative trait.

Information and entropy: Fisher saw "natural selection [as] a mechanism for generating an exceedingly high degree of improbability" (Edwards, 2000); qualitatively, we see natural selection as building up the

Accepted manuscript

ACCEPTED MANUSCRIPT

highly improbable combinations of alleles that cause reproductive success, and therefore, as being responsible for the functional information that is encoded in the genome. Fundamentally, the number of offspring left by an individual provides the information that, over very many generations, creates complex function. We conclude our paper by contrasting two distinct views as to how selection generates information.

Kimura (1961) pointed out that there is a close quantitative relation between Haldane's (1957) "cost of natural selection", and the increase in information due to selection (see also Worden, 1995). In order to pick out a single highly improbable genotype, which would have a chance $P = \prod_i p_i$ of fixing in the absence of selection, there must be at least $\log(1/P)$ selective deaths. More precisely, the difference in reproductive rate between the fittest genotype and the population average sums over time to a total of $\log(1/P)$. This is just the information added by choosing a single genotype that has *a priori* probability P.

Unfortunately, Kimura's (1961) argument has restricted scope. The "cost of selection" is precisely related to the increase in information if reproduction is asexual, but fails when there is sexual reproduction, and when genes interact with each other. To see this, note that truncation selection can raise any number of favourable alleles to moderate frequency for the same cost. If a fraction θ of the population is selected in each generation, then a large number of different favourable alleles, each initially rare, may all be selected in each generation, and can rapidly increase by a factor $\frac{1}{\theta}$ in each generation. Once they all become common, recombination is needed to bring them together into the fittest genotype, and so with asexual reproduction, progress stalls (Fisher, 1930; Muller, 1932). With recombination, however, the fittest genotype can rapidly fix, and most of the "cost of selection" is avoided: an arbitrarily improbable genotype can be fixed for a limited number of selective deaths.

There are several key differences between Kimura's (1961) argument, and the statistical entropy, S_H, which we have discussed here. Haldane's (1957) and Kimura's (1961) argument is based on a deterministic model of either asexual reproduction, or of strictly multiplicative gene effects; it relates the increase of information due to selection to the integral of mean fitness over time. In contrast, our results for the stochastic evolution of the distribution of allele frequencies assume free recombination, and apply for arbitrary gene interaction; we focus on the statistical balance between erosion of mean fitness by random drift, and its increase under selection. Nevertheless, on both views selection acts to pick out highly improbable genotypes, in much the same way that Maxwell's Demon can generate useful work from thermal energy by picking out particular molecules (Leff and Rex, 2003). We believe that the analogy with thermodynamics may lead to a better understanding of just how natural selection builds up the information that specifies complex organisms.

Acknowledgements

We are grateful to M. Cates, H.P. de Vladar and G. Sella for their helpful comments. This work was supported by a Royal Society/Wolfson Award, and by grants EP/T11753/01, EP/C546318/01 from the EPSRC.

Accepted manuscript

ACCEPTED MANUSCRIPT

References

Aita, T., and Y. Husimi, 2003. Thermodynamical interpretation of an adaptive walk on a Mt. Fuji-type fitness landscape: Einstein relation-like formula holds in a stochastic evolution. *Journal of Theoretical Biology* **225**, 215-228.

Aita, T., S. Morinaga, and Y. Husimi, 2005. Thermodynamical interpretation of evolutionary dynamics on a fitness landscape in an evolution reactor, II. *Bulletin of Mathematical Biology* **67**, 613-635.

Akin, E., 1979. The geometry of population genetics. *Lecture Notes in Biomathematics* **31.** Springer-Verlag. Berlin

Antonelli, P.L. and Strobeck, S., 1977. The geometry of random drift I Stochastic distance and diffusion. *Advances in Applied Probability* **9**, 238-249.

Ao, P., 2005. Laws in Darwinian evolutionary theory. *Physics of Life Reviews* **2**: 117-156.

Ao, P., 2008. Emerging of stochastic dynamical equalities and steady state thermodynamics from Darwinian dynamics. *Communications in Theoretical Physics* **49**: 1073-1090.

Baake, E., M. Baake, and H. Wagner, 1997. The quantum Ising chain is equivalent to a model of biological evolution. *Phys. Rev. Lett.* **78**, 559-562.

Baake, E., and H. Wagner, 2001. Mutation-selection models solved exactly with methods of statistical mechanics. *Genetical Research* **78**, 93-118.

Barton, N.H., 1989. The divergence of a polygenic system subject to stabilizing selection, mutation and drift. *Genetical Research* **54**, 59-77.

Barton, N.H. and Keightley, P.D., 2002. Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**, 11-21.

Barton, N.H. and S. Rouhani, 1987. The frequency of shifts between alternative equilibria. *Journal of Theoretical Biology* **125**, 397-418.

Barton, N.H. and S. Rouhani, 1993. Adaptation and the 'shifting balance'. *Genetical Research* **61**, 57-74.

Barton, N.H. and H.P. de Vladar, 2009. Statistical mechanics and the evolution of quantitative traits. *Genetics* doi: 10.1534/genetics.108.099309.

Berg, J., Willmann, S. and M. Lässig, 2004. Adaptive evolution of transcription factor binding sites. *BMC Evolutionary Biology* **4**, 42.

Boltzmann, L., 1872. Further studies on the thermal equilibrium of gas molecules. *Sitzungsberichte der Akademie der Wissenschaften, Wien* II **66**, 275-370. Reprinted in English in Brush, S. G. (1966*) Kinetic Theory Vol. 2 Irreversible processes*, pp. 88-175. Pergamon Press, London.

Boltzmann, L., 1877. Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respecktive den Sätzen über das Wärmegleichgewicht. *Wiener Berichte* **76**:373-435.

Bomze, I. M., 1991. Cross entropy minimization in uninvadable states of complex populations. *Journal of Mathematical Biology* **30**: 73-87.

Brooks, D.R. and E.O. Wiley, 1986. *Evolution as entropy*. Chicago University Press, Chicago.

Bulmer, M.G., 1980. *The mathematical theory of quantitative genetics*. Oxford University Press, New York.

Accepted manuscript

ACCEPTED MANUSCRIPT

Burch, C.L., Chao, L., 2000. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406**, 625-628.

Cercignani, C., 1998. *Ludwig Boltzmann: the man who trusted atoms*. Oxford University Press, Oxford.

Charlesworth, B., 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research* **63**, 213-228.

Coyne, J.A., Barton, N.H. and M. Turelli, 1997. A critique of Wright's shifting balance theory of evolution. *Evolution* **51**, 643-671.

Davis, B. K., 1994. On producing more complexity than entropy in replication. *Proceedings of the National Academy of Sciences of the United States of Americ*a **91**: 6639-6643.

De Visser, J.A.G.M., Hermisson, J., Wagner, G.P., Ancel-Meyers, L., Bagheri-Chaichian, H., Blanchard, J.L., Chao, L., Cheverud, J.M., Elena , S.F., Fontana, W., Gibson, G., Hansen, T.F., Krakauer, D., Lewontin, R.C., Ofria, C., Rice, S.H., Von Dassow, G., Wagner, A. and M.C. Whitlock, 2003. Evolution and detection of genetic robustness. *Evolution* **57**, 1959-1972.

Demetrius, L., 1997. Directionality principles in thermodynamics and evolution. *Proceedings of the National Academy of Sciences (U.S.A.)* **94**, 3491-3498.

Depew, D.J. and B.H. Weber, 1995. *Darwinism evolving*. MIT Press, Cambridge, Massachussetts.

Desai, M.M., Weissman, D. and M.W. Feldman, 2007. Evolution can favor antagonistic epistasis. *Genetics* **177**, 1001-1010.

Edwards, A,W,F., 2000. The Genetical Theory of Natural Selection. *Genetics* **154**, 1419-1426.

Eigen, M., 1971. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465-523.

Eigen, M., and P. Schuster, 1977. The hypercycle, a principle of natural selforganization A: Emergence of the hypercycle. *Naturwissenschaften* **64**, 541-565.

Fisher, R.A., 1930, *The genetical theory of natural selection*. Oxford University Press, Oxford.

Ginzburg, L.R., 1977. A macro-equation of natural selection. *Journal of Theoretical Biology* **67**, 677-686.

Haldane, J.B.S., 1957. The cost of natural selection. *Journal of Genetics* **55**, 511-524.

Hansen, T.F., 2006. The evolution of genetic architecture. *Annual Review of Ecology, Evolution, and Systematics* **37**, 123-157.

Hermisson, J., O. Redner, H. Wagner, and E. Baake, 2002. Mutation-selection balance: ancestry, load and maximum principle. *Theoretical Population Biology* **62** :9-46.

Iwasa, Y., 1988. Free fitness that always increases in evolution. *Journal of Theoretical Biology* **135**, 265-282.

Jaynes, E.T., 1983. *Papers on probability, statistics and statistical physics*. Kluwer Academic Publishers, Dordrecht.

Johnson, T. and N.H. Barton, 2002. The effect of deleterious alleles on adaptation in asexual populations. *Genetics* **162**, 395-411.

Keizer, J.L., 1987. *Statistical thermodynamics of nonequilibrium processes*. Springer Verlag, New York.

Kimura, M., 1961. Natural selection as the process of accumulating genetic information in adaptive evolution. *Genetical Research* **2**, 127-140.

Kimura, M. and T. Ohta, 1970. Genetic loads at a polymorphic locus maintained by frequency dependent selection. *Genetical Research* **16**, 145-150.

Accepted manuscript

ACCEPTED MANUSCRIPT

Kondrashov, A.S., 1995. Contamination of the genome by very slightly deleterious mutations, why have we not died 100 times over? *Journal of Theoretical Biology* **175**, 583-594.

Krakauer, D. C., and J. B. Plotkin. 2002. Redundancy, antiredundancy, and the robustness of genomes. Proceedings of the National Academy of Sciences (U.S.A.) 96: 9716-9720.

Kullback, S., 1987. The Kullback-Leibler distance. *The American Statistician* **41**, 340-341.

Kullback, S. and R.A. Leibler, 1951. On information and sufficiency. *Annals of Mathematical Statistics* **22**, 79-86.

Landau, L.D. and E.M. Lifshitz, 1980. *Statistical physics Part 1*. 3 rd edn. Butterworth –Heinemann. Oxford.

Le Bellac, M., F. Mortessagne, and G. G. Batrouni, 2004. *Equilibrium and non-equilibrium statistical thermodynamics*. Cambridge University Press, Cambridge.

Leff, H. S., and A. F. Rex, 2003. *Maxwell's Demon 2*. Institute of Physics, Bristol.

Lewontin, R.C., Moore, J.A., Provine, W.B. and B. Wallace, 1981. *The scientific work of Th. Dobzhansky*. pp. 93-114 *in* Lewontin RC, Moore JA, Provine WB, Wallace B, eds. *Dobzhansky's "Genetics of natural populations" I-XLIII*. Columbia University Press, New York.

Lotka, A.J., 1922. Natural selection as a physical principle. *Proceedings of the National Academy of Sciences (U.S.A.)* **8**, 151-154

Lynch, M. 2007. *The origins of genome architecture*. Sinauer Associates, Sunderland, Massachussetts.

Muller, H.J., 1932. Some genetic aspects of sex. *American Naturalist* **66**, 118-138.

Orr, H. A., 1998. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* **52**:935-949.

Prigogine, I., G. Nicolis and A. Babloyantz, 1972. Thermodynamics of evolution. *Physics Today* November, 23-28.

Provine, W., 1986. *Sewall Wright and evolutionary biology*. University of Chicago Press, Chicago

Rattray, M., Shapiro, J.L., 2001. Cumulant dynamics of a population under multiplicative selection, mutation, and drift. *Theoretical Population Biology* **60**, 17-31.

Rogers A. and A. Prugel-Bennett, 2000. Evolving populations with overlapping generations. *Theoretical Population Biology* **b**, 121-130.

Saakian, D. B., and C. K. Hu, 2004. Eigen model as a quantum spin chain: exact dynamics. *Physical Review E* **69**, 021913.

Schrodinger, E. 1944. *What is life ? The physical aspect of the living cell*. Cambridge University Press. Cambridge.

Schuster, P. and J. Swetina, 1988. Stationary mutant distributions and evolutionary optimization. *Bulletin of Mathematical Biology* **50**, 635-660.

Shahshahani, S., 1979. New mathematical framework for study of linkage and selection. *Memoirs of the American Mathematical Society* **17**

Sella, G. and A.E. Hirsh, 2005a. The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences (U.S.A.)* **102**, 9541-9546.

Sella, G. and A.E. Hirsh, 2005b. Correction to "The application of statistical physics to evolutionary biology". *Proceedings of the National Academy of Sciences (U.S.A.)* **102**, 14475-14475.

Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**, 379-423, 623-656.

Accepted manuscript

ACCEPTED MANUSCRIPT

Shapiro, J.L., Prügel-Bennett, A. and M. Rattray, 1994. A statistical mechanics formulation of the dynamics of genetic algorithms. *Lecture Notes in Computer Science* **865**, 17-27.

Turelli , M. and N.H. Barton, 2006. Will population bottlenecks and multilocus epistasis increase additive genetic variance? *Evolution* **60**, 1763-1776.

van Nimwegen, E., J. P. Crutchfield, and M. Huynen, 1999. Neutral evolution of mutational robustness. Proceedings of the National Academy of Sciences of the United States of America 96:9716-9720.

Wagner, A., 2005. *Robustness and evolvability in living systems*. Princeton University Press, Princeton, NJ.

Wicken, J., 1980. A thermodynamic theory of evolution. *Journal of Theoretical Biology* **87**, 9-23.

Wilke, C. O., 2005. Quasispecies theory in the context of population genetics. *BMC Evolutionary Biology* **5**: 44-51

Wilke, C. O., J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami, 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**: 331-333.

Wilson, A., 2008. Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems. *J. Roy. Soc. Interface* **5**, 865-871.

Worden, R.P., 1995. A speed limit for evolution. *Journal of Theoretical Biology* **176**, 137-152.

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

Accepted manuscript

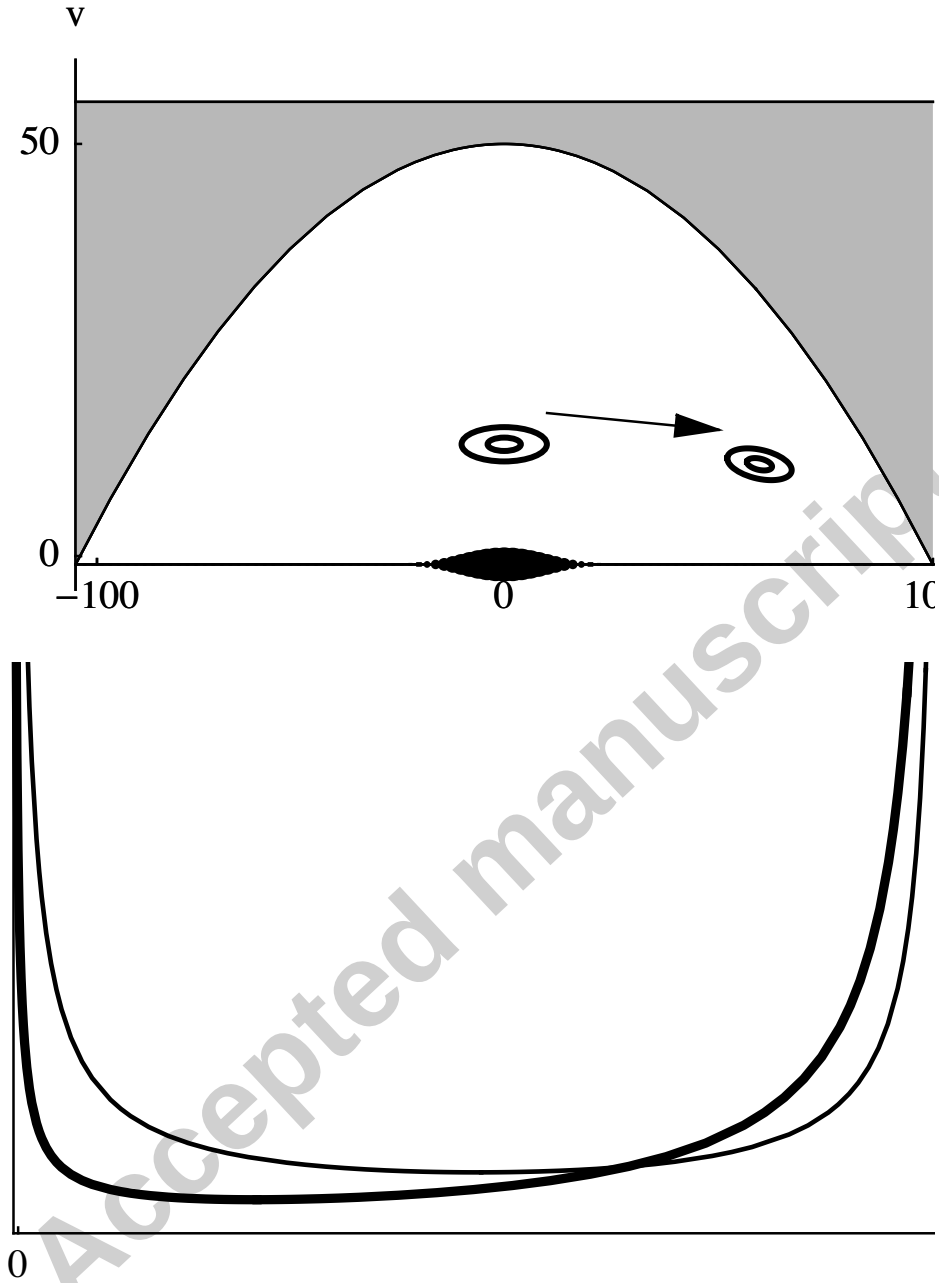
ACCEPTED MANUSCRIPT

Figures

Figure 1. The stationary distribution of the trait mean and variance is the product of three terms (Eq. 5): $P_0 \sim \overline{W}^{2N} e^{2\mathrm{NU}} e^{S_\Omega}$. Figure 1a illustrates the simple case of directional selection on an additive trait ($W = e^{sz}$). This shows the state of the population, described by the trait mean and variance ($\bar{z}$, v). With $n = 100$ loci of equal effect γ=1, $-n \le \bar{z} \le n$, $0 \le v \le \frac{n}{2}\left(1 - \left(\frac{\bar{z}}{n}\right)^2\right)$; the upper limit on the trait variance is shown by the parabola. First, suppose that there is no selection, and negligible mutation. Then, the entropy term e^{S_Ω} is dominated by fixed states ($p = 0,\ 1$). The number of states with given $\bar{z}$ follows *a* binomial distribution, and is dominated by states with intermediate mean. This is shown by dots on the horizontal axis ($v = 0$), around $\bar{z} = 0$: each state is marked by a disc with size proportional to the log number of states, S_W. Now, consider the combined effects of mutation and random drift, which are given by the product $e^{2NU} e^{S_\Omega}$. This is shown by the contours at upper middle (10%, 50% quartiles), for $2N\mu = 0.1$; the distribution is close to Gaussian. Finally, with selection $2N\beta = 1$, the distribution shifts to the right, indicated by the arrow, towards the state with maximum fitness ($\bar{z} = n$, $v = 0$). b) Even though the macroscopic variables are clustered closely around their expectations, the distribution of allele frequencies at individual loci is widely spread (thin line: no selection; thick line: $2N\beta = 1$; for both, $2N\mu$=0.1).

Accepted manuscript

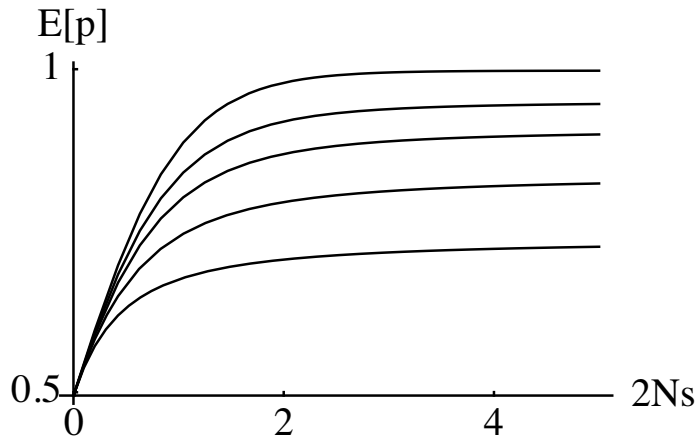
ACCEPTED MANUSCRIPT



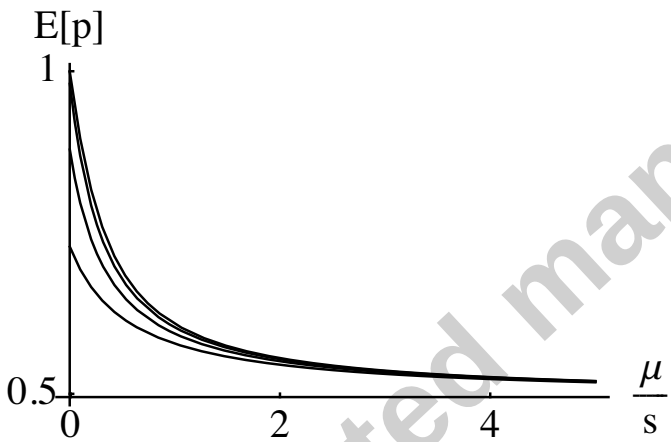
Accepted manuscript

ACCEPTED MANUSCRIPT

Figure 2. a) Mean allele frequency plotted against the scaled population size, $2Ns$, for $\frac{\mu}{s} = 0,\ 0.025,\ 0.05, 0.1, 0.2$ (top to bottom). b) Mean allele frequency plotted against $\frac{\mu}{s}$ for $2Ns$=0.5, 1, 2, 4 (bottom to top).



a

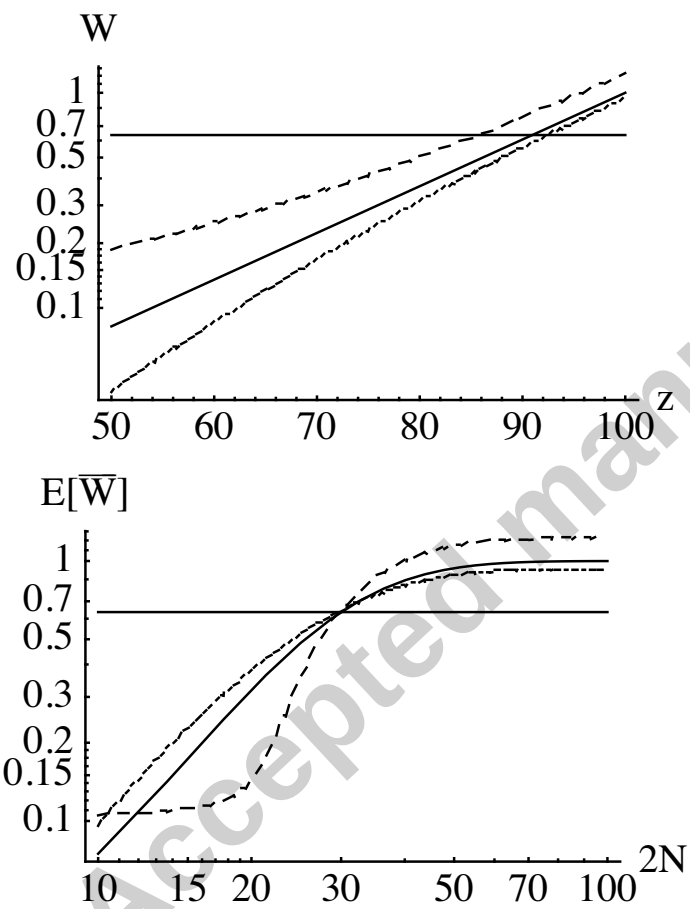


b

Accepted manuscript

ACCEPTED MANUSCRIPT

Figure 3. Selection acts on the sensitivity of fitness to an additive trait, z. Figure 3a compares four different relationships between fitness and trait, with the form $W = C * \exp\left(-b_1(n-z) + \frac{b_2}{2}(n-z)^2\right)$. The horizontal line represents insensitivity to the trait ($b_1 = b_2 = 0$). The solid straight line represents a mutliplicative dependence ($b_1 = 0.05,\ b_2 = 0$), and the upper and lower dashed lines represent synergistic ($b_1 = 0.05,\ b_2 = -0.0005$) and antagonistic (($b_1 = 0.05,\ b_2 = +0.0005$)) relations, respectively. Figure 3b shows how mean fitness depends on population size ($2N$), assuming low mutation rates ($N\mu << 1$). The overall fitness, C, is adjusted so that the mean fitnesses are equal at $2N = 30$. For larger population sizes, the relation that gives the genotype with the highest possible fitness is favoured (intercept at upper right of Fig. 3a; dashed line). In contrast, at lower population sizes, a flatter slope (b_1 small) and synergistic (negative) epistasis (b_2 negative; short dashed line) is favoured.



Accepted manuscript

ACCEPTED MANUSCRIPT

Supplementary Information

A: Relation with Aita et al. (2003, 2005)

Aita and Husimi (2003) analyse an "adaptive walk", by analogy with thermodynamics. They assume a sequence of length ν, with λ alleles at each site, and follow the change in a single sequence over successive time steps. At each time step, N offspring are produced, each of which differs by mutation at d sites; each mutation consists of a random choice amongst the $\lambda - 1$ alternative alleles at each of the d sites. The single sequence with highest fitness is then selected, and the process is repeated. Fitness is calculated as the sum of effects of the ν sites, with alleles having effects $0, \epsilon, \ldots (\lambda - 1)\epsilon$, where $\epsilon<0$. Thus, selection tends towards one fittest genotype, with fitness zero, but mutation and random sampling pull the walk into a stationary distribution with mean fitness below the adaptive peak.

This model represents the kind of optimisation algorithm that is used in evolutionary computation. However, it can also be seen as representing several biological models. The 'adaptive walk' across sequence space is similar to Iwasa's (1988) model of codon usage, and to the models of Sella and Hirsh (2005a). It is similar to Orr's (1998) analysis of Fisher's (1930) geometric model of phenotypic adaptation, in which fitness depends on the sum of squares of traits, and the probability of a jump to a new phenotype is proportional to its fitness advantage, so that fitness necessarily increases. Finally, if one thinks of an ensemble of adaptive walks, Aita and Husimi's (2003) model represents asexual evolution, with strong truncation selection within families, but no selection between families.

Aita and Husimi (2003) find the mean and variance of changes in fitness, W, by approximating the distribution of mutational effects as a truncated normal; this leads to approximations for the stationary distribution of fitness, and the typical time taken to reach this distribution. Aita and Husimi (2003) define a free fitness whose expecation always increases, as $G = W + TS$. Their entropy, S, is defined as being proportional to $\log(\Omega)$, where $\Omega[W]$ is the density of sequences with fitness W; this is essentially the same as our definition of S_Ω (Eq. 5). Their temperature, T, is chosen so that the mean fitness maximises G; it is roughly proportional to $\sqrt{d/\log(N)}$, Aita and Husimi (2003) go on to show that as the system approaches stationarity, the rate of increase in W is proportional to the gradient of free fitness; they divide this rate into two components, one due to selection for increasing W, and the other to evolution towards states with higher entropy.

In a subsequent paper, Aita et al. (2005) allow for a distribution of mutation distances, d, allow natural selection, by which M individuals are selected from N offspring, and use the NK model of epistasis, as well as the strictly additive case.

Although this analogy with thermodynamics is similar to ours, it is largely qualitative, and differs in several respects. The fluctuations are due to the combined action of sampling and mutation, and Aita and Husimi (2003, p. 216) see 'temperature' as depending on both processes. In contrast, our population genetic analysis treats mutation and selection as deterministic processes, which are separate from random sampling drift.

Accepted manuscript

ACCEPTED MANUSCRIPT

What is needed to apply our method to this model is a diffusion model for evolution over the sequence space (i.e., over microstates), rather than for the evolution of the single macroscopic variable, W. It is not clear that the stationary state would show detailed balance. If it did, it would be possible to write down a potential function, and hence the stationary distribution over sequences. The stationary distribution would then be proportional to $\exp(S_\Omega)$, as in Eq. 5 above; in our notation, fluctuations are proportional to the 'temperature', $1/2N$. Similarly, the expected value of the free fitness would necessarily increase as the population approches stationarity. However, there will not in general be a simple relation between the rate of change of the mean fitness, $dE[W]/dt$, and the gradient in free fitness, $\partial G/\partial W$. This question is investigated further by Barton and de Vladar (2009).

B: Drift load

Random drift reduces the mean fitness of a population below the maximum possible. Sella and Hirsh (2005a, Eq. 13) calculate this "fixed-drift" load for a specific model of mutation. If genotypes have a uniform distribution of fitnesses 0<W<1, and mutation increases or decreases fitness in a symmetrical way, then this load is $\frac{1}{2N}$ for $N >> 1$. This is twice the load due to random drift around a balanced polymorphism, which can be derived from Eq. 2 (assuming that selection is strong relative to mutation and drift; Kimura and Ohta, 1970). However, the drift load only takes this simple form with this particular choice of fitnesses: in general, it depends on selection as well as population size. For example, with symmetric mutation and $Ns >> 1$, the probability that a population is fixed for an allele which reduces fitness by s is $\sim \exp(-4Ns)$ in diploids, and so the fixed-drift load is $\sim 1 - s\, e^{-4Ns}$.

We note here that when fitness is reduced by mutation, the mean fitness of an *asexual* population depends only on the total mutation rate, and evolution of the relation between fitness and genotype then has no long-term effect on fitness. In contrast, the drift load is reduced by increased selection against deleterious alleles, because they are then less likely to fix by chance.

C: Alternative metrics

An alternative choice for a measure on the space of allele frequencies, midway between our and Iwasa's (1988) convention, is to define $U^* = 2\sum_{k=1}^{n}\left(\left(\mu_{P,k} - \frac{1}{8N}\right)\mathrm{Log}[p_k] + \left(\mu_{Q,k} - \frac{1}{8N}\right)\mathrm{Log}[q_k]\right)$, and to change the last term in Eq. 4 to $S_H = -E\left[\mathrm{Log}\left[P\left(\prod_{k=1}^{n}\sqrt{p_k q_k}\right)\right]\right]$. This choice is motivated by the fact that the space of allele frequencies has a natural metric (known as the Shashahani metric) which determines the effects of both drift and selection (Antonelli and Strobeck, 1977; Shahshahani, 1979; Akin, 1979, p. 37; Barton and Rouhani, 1987). Fluctuations in allele frequency are uncorrelated across loci (assuming linkage equilibrium), and have variance $\frac{p_i q_i}{2N}$; and the expected rate of change of allele frequency due to selection is $\frac{p_i q_i}{2}\partial_{p_i}\log[\overline{W}]$. Therefore, it is natural to measure the distance between states separated by $\delta\,\underline{p}$ as $\sum_{k=1}^{n}\left(\delta p_k^2/(p_k q_k)\right)$. Then, selection increases mean fitness as rapidly as possible, for a given rate of change of this measure, and random drift causes a uniform rate of diffusion. On this measure, the probability density is now $P^* = P\left(\prod_{k=1}^{n}\sqrt{p_k q_k}\right)$; the change in metric leads to an additional 'force', which we include (arbitrarily) by modifying the mutation potential. From a mathematical point of view, this choice is more elegant. However, we maintain a simpler biological interpretation in Eq. 4, by ensuring that the three terms in the free fitness depend solely on selection, mutation and drift, respectively, and that these are each proportional to the rates of selection, mutation, and drift $\left(s,\ \mu,\ \frac{1}{2N}\right)$.

Accepted manuscript

ACCEPTED MANUSCRIPT

D: Maximum entropy

Here, we show that the stationary distribution, P_0, maximises the entropy, S_H, subject to constraints on the expected values of a set of observables (see Le Bellac et al., 2004, p. 64, for a treatment in a physical context). We write the potential function in the form $\log(\overline{W}) + U = \vec{\alpha}.\vec{A}$, where the vector $\vec{A}$ is a function of the allele frequencies $\vec{p}$, and $\vec{\alpha}$ is a vector of coefficients. Crucially, the observables $\vec{A}$ may be a nonlinear function of the microscopic variables, $\vec{p}$. The simplest choice would be to set $\alpha_1 = \mu$, $\alpha_2 = s$, as measures of the rates of mutation and selection. Then, $A_1 = 2\sum_{k=1}^{n}(\theta_k \operatorname{Log}[p_k] + (1-\theta_k)\operatorname{Log}[q_k])$, where $\theta_k = \mu_{P,k}/(\mu_{Q,k} + \mu_{P,k})$, and $A_2 = \log(\overline{W})/s$ determines the form of selection. We might further separate $\log(\overline{W})$ into separate sources of selection: for example, with stabilising selection of strength s towards an optimum at z_{opt}, $\log(\overline{W}) = -s\frac{v}{2} - \frac{s(\bar{z}-z_{\text{opt}})^2}{2} = -s\frac{v}{2} - s\frac{\bar{z}^2}{2} + s\bar{z}z_{\text{opt}} - \text{constant}$. Thus, we can set $\vec{A} = \left\{2\sum_{k=1}^{n}(\theta_k \operatorname{Log}[p_k] + (1-\theta_k)\operatorname{Log}[q_k]), -\frac{v}{2}, -\frac{\bar{z}^2}{2}, \bar{z}\right\}$; the coefficients $\vec{\alpha} = \{\mu, s, s', sz_{\text{opt}}\}$ then represent mutation, selection to reduce variance in the trait, v, stabilising selection to reduce deviations in the mean, $\bar{z}^2$, and directional selection on the trait mean, $\bar{z}$. We might also add observables that do not affect fitness, but are nevertheless of interest, by setting their α_k to zero.

Generalising the definition of S_H given below Eq. 4 we write:

$$S_H[P] \equiv \int P \log\left(\frac{\phi}{P}\right) d\vec{p} \tag{6}$$

where ϕ is a measure which we take here to be $\phi = \prod_{k=1}^{n}(p_k q_k)^{-1}$. To find the distribution P_{ME} that maximises S_H subject to constraints on the expectations, $\langle\vec{A}\rangle$, we use the method of Lagrange multipliers, setting these multipliers to be proportional to $2N\vec{\alpha}$. We also require the constraint that $\int P_{\text{ME}}\, d\vec{p} = 1$, with associated multiplier denoted by $2N\gamma$:

$$\begin{aligned} 0 &= \delta S_H + 2N\gamma\delta\left(\int P\, d\vec{p}\right) + 2N\vec{\alpha}.\delta\langle\vec{A}\rangle \\ &= \int\left(\log\left(\frac{\phi}{P}\right) - 1\right)\delta P\, d\vec{p} + 2N\gamma\int\delta P\, d\vec{p} + 2N\int\vec{\alpha}.\vec{A}\,\delta P\, d\vec{p} \\ &= \int\left(\log\left(\frac{\phi}{P}\right) + (2N\gamma - 1) + 2N\vec{\alpha}.\vec{A}\right)\delta P\, d\vec{p} \end{aligned} \tag{7}$$

Rewriting the normalisation as $Z = \exp(1 - 2N\gamma)$, we find that the distribution P_{ME} that maximises S_H, for given values of $\langle\vec{A}\rangle$, is:

$$P = \frac{1}{Z}\phi\, e^{2N\vec{\alpha}.\vec{A}} \text{ where } Z = \int\phi e^{2N\vec{\alpha}.\vec{A}}\, d\vec{p} \tag{8}$$

The coefficients $\vec{\alpha}$ determine the values of the expectations through the constraint:

$$\langle\vec{A}\rangle = \frac{1}{Z}\int\phi\,\vec{A}\, e^{2N\vec{\alpha}.\vec{A}}\, d\vec{p} \tag{9}$$

The expectations can also be found by differentiating the normalisation:

Accepted manuscript

ACCEPTED MANUSCRIPT

$$< \vec{A} > = \frac{1}{2N} \frac{\partial \log(Z)}{\partial \vec{\alpha}} \tag{10}$$

(c.f. Le Bellac et al. 2004, Eq. 2.66). Moreover, the covariance amongst the observables is given differentiating again:

$$\mathrm{cov}\left(A_i, A_j\right) = C_{i,j} = \frac{1}{4N^2} \frac{\partial^2 \log(Z)}{\partial \alpha_i \, \partial \alpha_j} \tag{11}$$

(c.f. Le Bellac et al. 2004, Eq. 2.70).

E: Relation between entropy measures

The entropies S_W, S_H are distinct, and indeed are functions of different variables: $S_\Omega[\vec{A}]$ is a functon of the observables, whereas $S_H[P]$ is a functional of the distribution across microstates, and does not depend on any definition of the observables. However, we show here that there is nevertheless a close relation between them.

Generalising the definition given below Eq. 5, we define $S_\Omega[\vec{A}]$ as the log density of states that are consistent with macroscopic variables $\vec{A}$:

$$S_\Omega[\vec{A}] \equiv \mathrm{Log}\left(\int_{\vec{A}} \phi \, d\vec{p} \right) \tag{12}$$

where $\phi[\vec{p}] = \left(\prod_{i=1}^{n} p_i \, q_i \right)^{-1}$ is a measure on the allele frequency space, and where the integral is over states $d\vec{p}$ that give macroscopic values in $d\vec{A}$. If ϕ were constant, then $S_\Omega[\vec{A}]$ would be the same as Boltzmann's entropy, as in the case of physical systems where all the micro-states consistent with the observables have the same probability (Landau and Lifshitz, 1980, p. 25). Assuming that the evolutionary forces depend only on $\vec{A}$, the stationary distribution of the observables, $P_0^*[\vec{A}]$ is then obtained by integrating over the distribution of allele frequencies, $P_0[\vec{p}]$, conditional on $\vec{A}$:

$$P_0^*\left(\vec{A}\right) = \frac{1}{Z} e^{2N\vec{\alpha}.\vec{A}} e^{S_\Omega} \tag{13}$$

What is the relation between $S_\Omega[\vec{A}]$ and $S_H[P]$? The maximum value of $S_H[P]$, given constraints on $\langle \vec{A} \rangle$, is found at the stationary density (Supplementary Information B). Substituting from Eq. 8:

$$S_{\mathrm{ME}}\left[\langle \vec{A} \rangle\right] = \int P_0 \, \mathrm{Log}\left(\frac{\phi}{P_0} \right) d\vec{p} = \mathrm{Log}(Z) - 2N\vec{\alpha}.\langle \vec{A} \rangle \tag{14}$$

(c.f. Le Bellac et al. 2004, Eq. 2.65). The normalisation factor Z is given by the requirement that $P_0^*\left(\vec{A}\right)$ integrates to 1 (Eq. 13). Thus:

$$S_{\mathrm{ME}}\left[\langle \vec{A} \rangle\right] = \mathrm{Log}\left(\int e^{2N\vec{\alpha}.\vec{A}} e^{S_\Omega} \, d\vec{A} \right) - 2N\vec{\alpha}.\langle \vec{A} \rangle \tag{15}$$

Writing $\vec{A}$ in terms of its deviation from expectation, $< \vec{A} > + \delta \vec{A}$, and similarly for the entropy, $S_\Omega[\vec{A}] = S_\Omega\left[< \vec{A} > \right] + \delta S_\Omega$:

Accepted manuscript

ACCEPTED MANUSCRIPT

$$S_{\mathrm{ME}}\left[\langle \vec{\mathrm{A}} \rangle\right] = S_{\Omega}\left[\langle \vec{\mathrm{A}} \rangle\right] + \mathrm{Log}\left(\int e^{2N\vec{\alpha}.\delta\vec{A}}\, e^{\delta S_{\Omega}}\, d\delta\vec{A}\right) \tag{16}$$

If $\vec{A}$ is clustered around its expectation, in an approximately Gaussian distribution, the integral can be evaluated to give:

$$S_{\mathrm{ME}} = S_{\Omega} - \frac{1}{2}\,\mathrm{Log}\left(\left\|\frac{S''_{\Omega}}{2\pi}\right\|\right) \tag{17}$$

where $\|\ldots\|$ denotes a determinant, and S''_{Ω} denotes the matrix of second derivatives of S_{Ω} w.r.t. $\vec{A}$, evaluated at $\langle\vec{A}\rangle$. Now, we know that the covariance of fluctuations in $\vec{A}$ is given by the matrix C (Eq. 11); hence, $2\,N S''_{\Omega} = C^{-1}$. Hence:

$$S_{ME} = S_{\Omega} + \frac{1}{2}\,\mathrm{Log}(\|2\,\pi C\|) \tag{18}$$

Thus, S_{ME} and S_{Ω} converge when n is large, so that fluctuations are small: then, S_{ME}, S_{Ω} are large compared with the logarithmic term.

Accepted manuscript