Genet. Res., Camb. (1995), **66**, *pp.* 241–253

The constraints of finite size in asexual populations and the rate of the ratchet

DAMIAN D. G. GESSLER
Biology Board of Studies, University of California, Santa Cruz, CA 95064, U.S.A. Phone: (408) 459-5147, Fax: (408) 459-4882. Email: gessler@biology.ucsc.edu

(*Received 23 September 1994 and in revised form 20 June 1995*)

Summary

An analysis of mutation accumulation in finite, asexual populations shows that by modeling discrete individuals, a necessary condition for mutation–selection balance is often not met. It is found that over a wide parameter range (whenever $N e^{-\mu/s} < 1$, where N is the population size, μ is the genome-wide mutation rate, and s is the realized strength of selection), asexual populations will fail to achieve mutation–selection balance. This is specifically because the steady-state strength of selection on the best individuals is too weak to counter mutation pressure. The discrete nature of individuals means that if the equilibrium level of mutation and selection is such that less than one individual is expected in a class, then equilibration towards this level acts to remove the class. When applied to the classes with the fewest mutations, this drives mutation accumulation. This drive is in addition to the well-known identification of the stochastic loss of the best class as a mechanism for Muller's ratchet. Quantification of this process explains why the distribution of the number of mutations per individual can be markedly hypodispersed compared to the Poisson expectation. The actual distribution, when corrected for stochasticity between the best class and the mean, is akin to a shifted negative binomial. The parameterization of the distribution allows for an approximation for the rate of Muller's ratchet when $N e^{-\mu/s} < 1$. The analysis is extended to the case of variable selection coefficients where incoming mutations assume a distribution of deleterious effects. Under this condition, asexual populations accumulate mutations faster, yet may be able to survive longer, than previously estimated.

1. Introduction

Muller (1964) noted the opportunity for a ratchet mechanism to operate in asexual populations. He noted that if those individuals with the fewest number of mutations fail to reproduce, or produce offspring with more mutations than they themselves have, then in the absence of back mutation or recombination, the minimum number of mutations per individual in the following generation will increase. Felsenstein (1974), in a paper identifying the importance of this process *vis-à-vis* the maintenance of recombination, named this monotonic accumulation of mutations in asexual populations Muller's ratchet.

The role of Muller's ratchet was further discussed by Maynard-Smith (1978), and in the same year Haigh (1978) published a mathematical model describing the process. Kimura & Maruyama (1966) had shown that the distribution of the number of mutations per individual is Poisson with mean μ/s, where μ is the genome-wide mutation rate and s is the invariant selection coefficient. Haigh (1978) proved that distribution is a unique, equilibrium distribution, and thus in the absence of perturbations, asexual populations will achieve mutation–selection balance: the only way mutations can accumulate past the equilibrium level (i.e. the only way the ratchet can operate), is by the stochastic loss of the best class. Although Muller never explicitly specified the mechanisms responsible for driving the ratchet, Haigh's analysis demonstrated the sole importance of sampling error. This conceptualization has become so established that subsequent authors have often defined the ratchet in terms of the stochastic loss of the best class. Previous authors have discussed the role of non-stochastic processes in Muller's ratchet (Haigh, 1978; Bell, 1982; Pamilo *et al.* 1987; Lynch & Gabriel, 1990; Charlesworth, D. *et al.* 1993; Gabriel *et al.* 1993; Stephan *et al.* 1993), yet most investigations into the rate of the ratchet have concentrated on quantifying the stochastic component.

The rate of the ratchet has remained essentially

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

unsolved for over thirty years: there is still no general solution. This paper demonstrates that many asexual populations will never achieve Haigh's equilibrium condition, and therefore mutations will accumulate for reasons in addition to the stochastic loss of the best class. This conclusion leads to an approximation of the rate of the ratchet over a parameter range that is important for metazoans and metaphyta.

2. The model

Monte Carlo simulations are used to model the accumulation of mutations in asexual (parthenogenic) individuals. Prior to starting a simulation, N_A individuals are initialized as mutation-free. The sequence of events in discrete generations is reproduction → mutation → selection, with the population (of offspring) being censused after mutation but prior to selection.

Each generation an individual is randomly picked with replacement from the pool of N_A adults. An offspring is cloned, the mutation process is applied to its genome, and a fitness-value is calculated. The fitness-value for each offspring is the multiplicative effect of all mutations in its genome, i.e. $\prod_{i=1}^{n}(1-s_i)$, where s_i is the selection coefficient of the ith allele in a genome of n mutations. This value is then compared with a uniform random variate, and if it is greater, the offspring is saved in a separate pool to become an adult in the next generation. This process is repeated until N_A new individuals are selected.

Mutation is modeled as a Poisson process of mean μ, where μ is the genome-wide mutation rate (Kimura & Maruyama, 1966). For modeling all mutations with the same deleterious effect, s is held constant (herein called constant s runs). When examining the consequence of variation in mutational effects, each mutation has a selection coefficient drawn from an approximate negative exponential distribution (Ohta, 1977; Gillespie, 1991) (herein called variable s runs). To allow the use of an analytical probability density function (p.d.f.) with unit area on the domain [0, 1], a beta distribution with shape parameters $\nu = 1$ and $\omega = s^{-1} - 1$ is used to approximate a negative exponential of mean s.

Before any statistics are recorded, mutations are allowed to accumulate until the mean population fitness first reaches its infinite-size expectation of $e^{-\mu}$ (Haldane, 1937; Muller, 1950; Kimura *et al.* 1963). Let **N** be a random variable that stands for the actual number of offspring selected to produce N_A adults in any one generation, and let $N = \mathrm{E}(\mathbf{N})$. Then on average, the number of offspring N needed to maintain a carrying capacity of N_A adults is $N = N_A \bar{w}^{-1} = N_A e^{\mu}$. To maintain this constant relationship between the number of individuals before selection (N) and the number after selection (N_A), offspring in subsequent generations are compared with a standard that is decremented proportional to the decay in mean fitness. For example, when mean fitness is $\bar{w} = e^{-1} \cong 0{\cdot}37$, offspring are compared with a $(0, 1 \times 0{\cdot}37^{-1} \times 0{\cdot}37) = (0, 1)$ uniform variate. As mean fitness falls to 0·36, offspring are compared with a $(0, 1 \times 0{\cdot}37^{-1} \times 0{\cdot}36) = (0, 0{\cdot}97)$ uniform variate, etc. By decreasing the standard of comparison as mutations accumulate, N is kept constant. Because N_A and N are independent of s, sampling error in both the adult and the offspring stage is the same in both constant and variable s simulations throughout the length of each simulation. The constancy of N_A and N assures that population extinction is impossible, making it conceptually distinct from the mutational melt-down approach of Lynch & Gabriel (1990). The general behavior of the model is similar to the models of Kimura & Maruyama (1966), Felsenstein (1974), Haigh (1978) and Charlesworth (1990) when parameterized in equivalent ways. In comparisons with Haigh's model, it is noteworthy that by changing the pool of individuals that selection acts on from that of N_A adults in his model to that of their offspring in this model, the relevant population size is increased by e^{μ}.

3. Mutation accumulation beyond the stochastic loss of the best class

Our current conception is that in the absence of stochasticity, asexual populations will achieve mutation–selection balance (Haigh, 1978; Higgs, 1994). But the conditions for equilibrium require the presence of individuals that may have a very low probability of actually being present. If they are

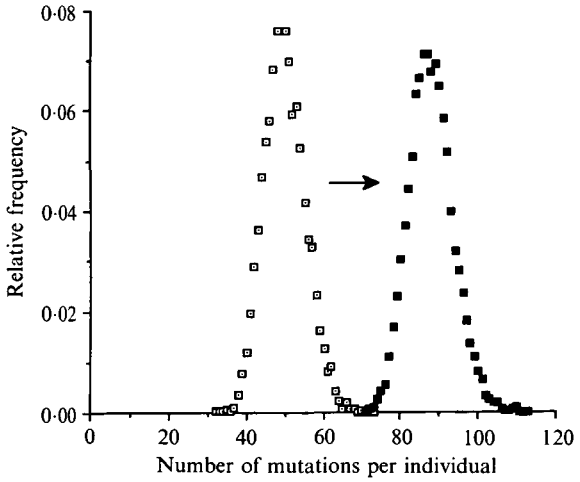

Fig. 1. In asexual populations, the relative selective advantage of each mutational class can be at its equilibrium (i.e. the distribution is at its equilibrium shape), but the population can still acquire mutations for reasons other than the stochastic loss of the best class. The rate the best individuals are lost is a combination of mutation pressure diminishing the best class, mutation-free reproduction augmenting the class, and the stochastic exclusion of some individuals from reproducing at all. The figure shows two snapshots of the same population taken at generation zero and generation 100. $N \cong 10^4$; $\mu = 1$; $s = 0{\cdot}02$.

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

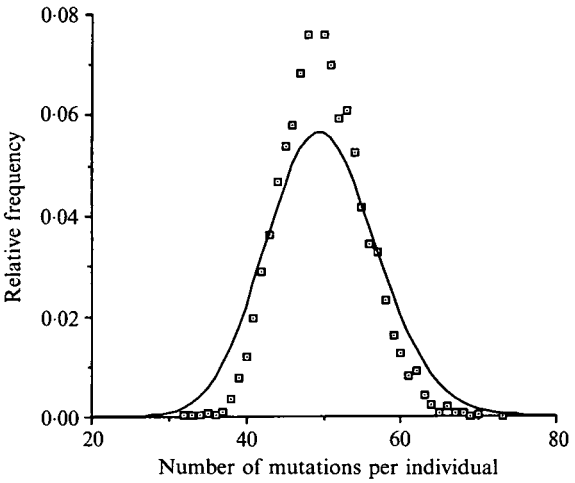

Fig. 2. Continuous-line representation of the discrete probability mass function (Poisson $[\theta = \mu/s; x]$) as it fits the simulation in Fig. 1 at generation zero. The scale on the abscissa has been increased to emphasize that the Poisson distribution overestimates the observed variance by 61%.

absent, the reduced range in fitness between the most fit individuals and the mean is insufficient to maintain enough variance in the population, and the population will be unable to maintain its equilibrium number of mutations (Fig. 1).

Because individuals are discrete, non-divisible units, whenever the equilibrium expectation of a class (i.e. the number of individuals) is less than unity, at some point the class will have to have zero individuals. For classes in the right-hand tail of the distribution, (or in both tails of a sexual population), lost classes can be rebuilt at a later time, so any non-zero probability mass can be achieved as a time-average. There is no restriction for the equilibrium number of individuals in these classes to be greater than unity. But for the best class in an asexual population, there is no regenerative mechanism once the class is lost. Thus if the equilibrium number of individuals in the class is less than unity, as mutation and selection equilibrate to a level less than one individual, they act to remove the individual and consequently drive the ratchet.

Consider a simulation parameterized with values similar to those found in the *Drosophila* literature (Fig. 2; $\mu = 1$ and $s = 0{\cdot}02$; Simmons & Crow, 1977; Crow & Simmons, 1983; Charlesworth *et al.* 1990; Crow, 1993*a*, *b*), where it is noted that the actual distribution of the number of mutations per individual is markedly hypodispersed compared to the Poisson expectation. As Haigh (1978) indicated, as mutations begin to accumulate, the entire distribution will march to the right, with both the mean and the best class moving at the same rate.

To the degree that the expected number of mutation-free individuals (the zero class) is below unity, it is proportionally unlikely to be realized in a finite population. For example, for the values used in Fig. 1 and 2 ($N \cong 10^4, \mu = 1, s = 0{\cdot}02, \theta = \mu/s = 50$), the expected size of the zero class is $N e^{-\theta} = 1{\cdot}9 \times 10^{-18}$ of an individual (eqn 6, Haigh, 1978). Even if an individual is placed in this class, the existence of this one individual exceeds the class' equilibrium size, the class is unstable, and there is deterministic pressure to remove this one individual. This is true not only for the zero class, but for all classes up to the first class with an expectation greater than or equal to unity. Although mutation pressure is a stochastic process, as a mechanism of evolution it is a non-stochastic pressure, and thus the inevitable removal of these classes is due, at least in part, to a force distinct from sampling error.

If the less mutated classes are missing, then one would perhaps expect that the mean (θ) could just shift to the right, thereby restoring the necessary range between the best class and the mean and establishing an equilibrium. But when the mean number of mutations is θ, the mean relative strength of selection on each class is at its steady-state magnitude and will remain time invariant (Haigh, 1978). Thus as the mean moves to the right, the equilibrium size for each class is also redefined. After the mean has moved one class, the first class has the same expectation that the zero class used to have, and thus it too will be lost in a similar manner. This interdependency of the stability of the best class and the mean on the range between them maintains the reduced range of the distribution. The rate the best class is lost is both a function of the probability of its stochastic loss, and also of the strength of selection on maintaining it under mutation pressure.

At equilibrium, each class maintains its size by a balance between mutation pressure diminishing it, mutation-free reproduction within the class augmenting it, and new mutants from less-mutated classes also augmenting it. Only the zero class can maintain its size solely by the first two mechanisms: all others require recruitment from less mutated classes. If these classes are unstable, then their eventual absence will reduce the recruitment rate into classes downstream, and they too will be unable to maintain their position. Because the steady-state strength of selection on all classes other than the zero class is below that required to offset mutation pressure, without a stable zero class, i.e. whenever $N e^{-\theta} < 1$, mutations will accumulate indefinitely.

It is true that for any θ there exists a N large enough to guarantee a stable zero class, but the required size of N grows exponentially with θ. For any N that does not satisfy $N e^{-\theta} \geqslant 1$, mutation pressure can be expected to reduce the number in the best class to below one individual. Despite this, the best class never has an expectation of exactly zero; therefore the continued existence of the best class when its expectation is less than unity reflects the action of stochasticity in maintaining the class above its expectation, while its final loss must be due to at least some stochasticity reducing the number below its

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

expectation. Thus some stochasticity is always required to drive the ratchet, though its magnitude may be negligibly small. In Figs. 1 and 2, the contribution of stochasticity (the time to lose $1{\cdot}9 \times 10^{-18}$ of an individual) is so small that the ratchet is driven primarily by mutation pressure. In this paper, the phrase 'non-stochastic component' refers to the observation that whenever $N e^{-\theta} < 1$, the Poisson equilibrium condition of mutation–selection balance is not only unattainable, but that a necessary consequence of its approach is the drive of the ratchet.

4. The distribution of the number of mutations per individual

The identification that asexual populations may fail to achieve mutation–selection balance because of the discrete nature of individuals, allows one to formulate a probability mass function (p.m.f.) for the distribution of the number of mutations per individual.

If the first k classes have a low expectation and are consequently absent, then when the distribution is at its equilibrium shape all non-zero probability associated with the remaining classes is now constrained to at most classes $k \ldots \infty$. With almost unit probability, all classes $0 \ldots k-1$ will be zero after some finite amount of time, even after adjusting for movement of the mean. The mutation-selection process is still identical to that analysed by Haigh (1978), so the equilibrium condition of a Poisson relationship among classes still holds, though the distribution is now a shifted Poisson distribution of parameter $\lambda = \theta - k$, where k is the first class with an expectation greater than or equal to unity

$$k = \min\{x \mid N e^{-\theta}\theta^x/x! \geqslant 1, x \in \{0, 1, 2, \ldots, \theta\}\}.$$

The shifting of the first class from the zero class to the kth class, while the mean still stays at θ, yields a p.m.f. for the distribution of the number of mutations per individuals as

$$\Pr(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{(x-k)}}{(x-k)!} & x \geqslant k \\ 0 & \text{otherwise} \end{cases} . \qquad (1)$$

Equation (1) identifies a shifted Poisson distribution, yet there are two factors in Monte Carlo simulations that act to change the observed distribution from that of (eqn 1). The first is that in this new distribution there also may be classes that have a low expectation, and thus they too may not be realized. The second is that the distance from k to the mean, and more importantly the distance from the best class to the mean, is a random variable. Variance in this distance is equivalent to variance in the parameter of the distribution, and this acts to increase the variance in the number of mutations per individual beyond that predicted by (eqn 1). Quantitatively, these points can produce a distribution markedly different from (eqn 1), yet qualitatively, the major conceptual distinction between populations where $N e^{-\theta} \geqslant 1$ and those where $N e^{-\theta} < 1$ is captured in (eqn 1) above. The remainder of this section incorporates the above points into a revised p.m.f.

To address the first point, let b equal the expected position of the best class. The effect of the missing classes $k \ldots b-1$ is not analogous to the missing $0 \ldots k-1$ classes. The missing classes $k \ldots b-1$ represent a left truncation on an otherwise (shifted) Poisson distribution, while the missing $0 \ldots k-1$ classes far exceed that of a left truncation: they substantially change the shape of the distribution from one of parameter θ to one of parameter $\lambda = \theta - k$ shifted k classes to the right. To include the effect of truncation, (eqn 1) may be refined into a truncated Poisson distribution shifted k classes to the right with the first non-zero probability at class b (cf. Haight, 1967; Johnson *et al.* 1993; see also the [left] displaced Poisson distribution of Staff, 1964, 1967)

$$\mathrm{f}(b; x) = \Pr(X = x) = \begin{cases} \dfrac{\lambda^{(x-k)}}{(x-k)!}\left[\displaystyle\sum_{i=b-k}^{\infty} \frac{\lambda^i}{i!}\right]^{-1} & x \geqslant b \\ 0 & \text{otherwise} \end{cases} . \qquad (2)$$

Equation (2) describes the distribution of the number of mutations per individual if the best class is actually lost whenever its frequency falls below one individual. This can be verified by constructing a new model that reiterates the mutation and selection operations on p.m.f.s (instead of stochastically picking individuals from a population) and truncates the best class accordingly (data not shown). Our interest, though, is in describing the more relevant case where there is a probabilistic loss of the best class.

If the mean number of mutations is large ($\mu/s > 10$) such that deviations around the relative position of the best class in reference to the mean are approximately symmetric, then the distance between the best class and the mean predicted by (eqn 2) will be close to the mean observed distance, and (eqn 2) can be used to approximate the position of the best class. If b equals the mean position of the best class, then b is the first class with an expectation of at least one individual

$$b = \min\{x \mid N\mathrm{f}(x; x) \geqslant 1, x \in \{k, k+1, k+2, \ldots, \theta\}\}. \qquad (3)$$

It is understood that to determine b initially, b in (eqn 2) is redefined for each $b = x$ in (eqn 3) until (eqn 3) is satisfied. The use of a truncated distribution allows for the extremely accurate estimations of the position of the best class shown in Table 1.

To address the second point, note that the distance between the best class and the mean affects the relative strength of selection on the best class, and thus variance in this distance is paramount to variance in the parameter of the distribution. If the parameter itself a random variable, then the resultant distribution

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

Table 1. *Simulation results over a variety of parameter values.* k *is the first class in a Poisson distribution of parameter* θ *that has an expectation greater than one individual.* b *is the first class in a shifted Poisson distribution of parameter* λ *that has an expectation greater than one individual. Each row represents the mean of 10 independent simulations, each run for* N *generations. Data was reported every* N/*100 generations, from which only the last* N/*2 generations were used. Missing entries assume the value in the above row*

				Best class		Worst class	
N	s	μ	k	Predicted b	Observed	Predicted	Observed
500	0·005	0·5	77	88	88·32	116	116·41
		1	168	186	185·72	219	218·78
		2	358	384	384·26	421	420·67
	0·01	0·5	33	40	40·38	64	64·14
		1	77	88	87·87	116	116·67
		2	168	186	186·08	219	218·33
	0·02	0·5	13	17	17·28	37	36·80
		1	33	40	39·70	64	64·87
		2	77	88	87·98	116	116·02
1000	0·005	0·5	74	86	86·68	119	118·68
		1	165	184	183·41	222	222·86
		2	352	382	381·17	426	425·10
	0·01	0·5	31	39	38·89	66	66·11
		1	74	86	86·06	119	119·47
		2	165	184	183·85	222	222·40
	0·02	0·5	12	16	16·27	38	38·34
		1	31	39	38·69	66	66·62
		2	74	86	86·42	119	118·88

	Variance			Rate of the ratchet	
$t = b-k$	Poisson $\theta = \mu/s$	Predicted $\mu/s-k+t$	Observed $\sigma_n^2 \pm$ S.E.	Predicted $R = sk-st$	Observed $R_{obs} \pm$ S.E. $\times 10^2$
11	100	34	33·24 ± 6·61	0·33	0·33 ± 3·15
18	200	50	47·00 ± 5·85	0·75	0·77 ± 4·18
26	400	68	65·65 ± 13·31	1·66	1·69 ± 7·87
7	50	24	21·32 ± 2·14	0·26	0·29 ± 2·33
11	100	34	32·87 ± 3·04	0·66	0·66 ± 4·58
18	200	50	47·79 ± 4·05	1·50	1·50 ± 6·00
4	25	16	13·92 ± 0·99	0·18	0·23 ± 1·45
7	50	24	23·40 ± 1·30	0·52	0·54 ± 3·67
11	100	34	32·17 ± 3·62	1·32	1·37 ± 5·62
12	100	38	35·29 ± 3·75	0·31	0·33 ± 1·51
19	200	54	56·21 ± 6·53	0·73	0·71 ± 3·19
30	400	78	81·98 ± 7·96	1·61	1·60 ± 5·58
8	50	27	24·03 ± 1·32	0·23	0·26 ± 1·33
12	100	38	38·68 ± 2·63	0·62	0·63 ± 2·43
19	200	54	55·77 ± 2·99	1·46	1·45 ± 4·71
4	25	17	15·25 ± 0·58	0·16	0·19 ± 1·34
8	50	27	24·49 ± 1·32	0·46	0·51 ± 2·21
12	100	38	36·79 ± 1·60	1·24	1·28 ± 2·70

is said to be a compound or mixed distribution. Teicher (1960) showed that a mixed Poisson distribution is itself never Poisson (except in the trivial case when the parameter is distributed as a causal [or deterministic] variate), and that a mixture on two mixed Poissons (i.e. through the mutation and selection process) remains a mixed Poisson. Thus our motivation is to determine a new distribution that accounts for the increased variance in the population accorded by stochasticity in the distance between the best class and the mean.

Figure 3*a*, *b* illustrates that although the best class and the mean are highly correlated, random fluctuations in the difference between the two can be appreciable. To quantify this variance, it will be convenient to translate the distribution k units to the left, and thus anchor the distribution at zero. This will allow derivations in terms of centralized, instead of

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

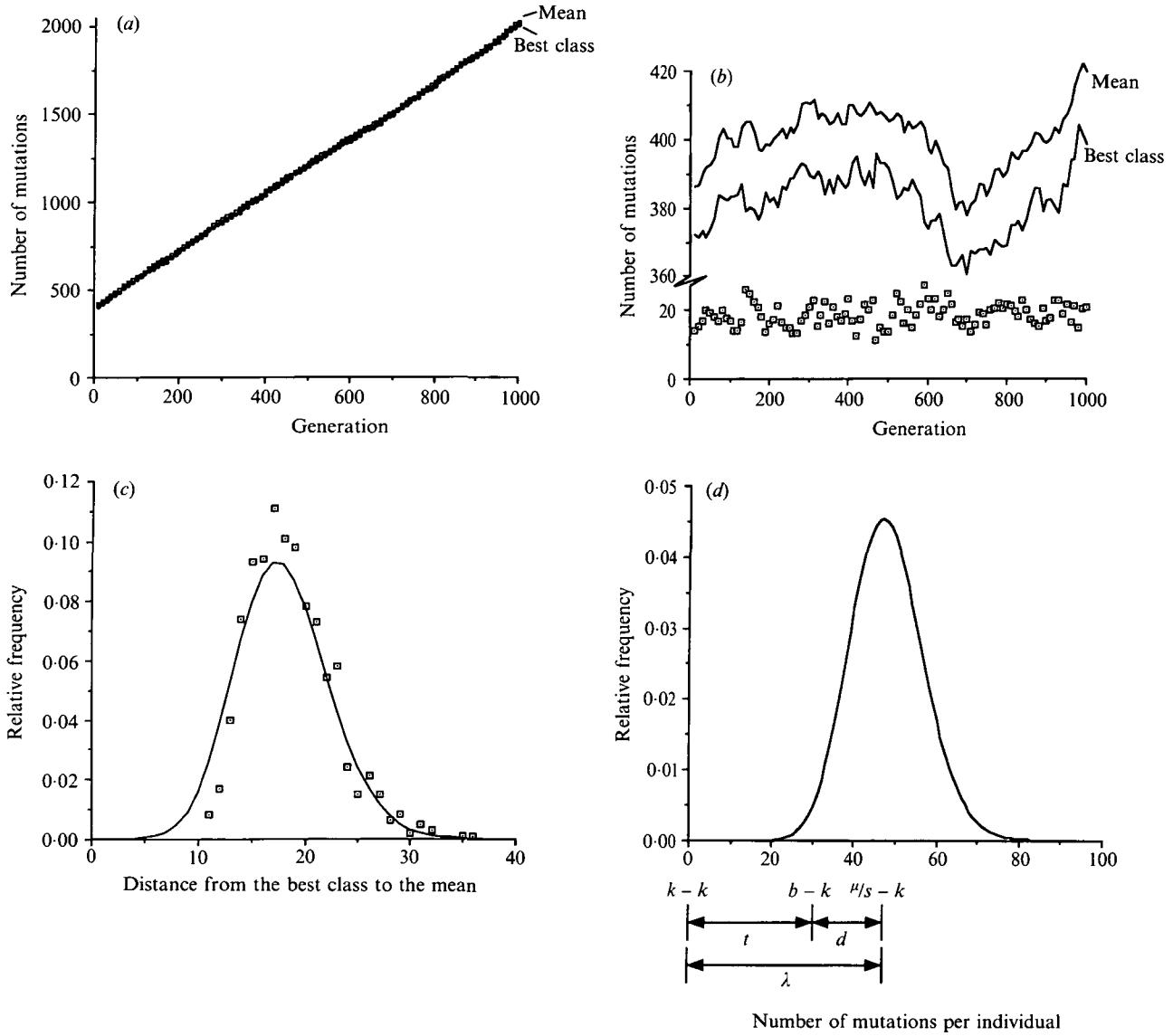

Fig. 3. (*a*) As mutations accumulate, the best class is continually redefined, so there always exists a non-zero best class. When the best class is lost, the second best class will tend to be larger than its expected size as the new best class. Because of the high correlation between the position of the best class and the mean, a correlated shift in the position of the mean maintains the average distance between the two classes. Thus variance in the distance between the best class and the mean tends to increase the average size of the best class (not shown on graph), and this slows the ratchet. (*b*) The mean number of mutations per individual and the number of mutations in the best class from (*a*). The continuous-line representations for the best class and the mean are plotted after adjusting for the rate of mutation accumulation. Adjustment is performed by subtracting a linear regression on the mean from the observed values. The data points are the observed distance between the mean and the best class. (*c*) The distribution of the distance from the best class to the mean as it fits a gamma distribution. Parameterization is defined subsequently in the text. To gain sufficient precision, each point represents the height of a histogram cumulative over ten independent runs. The bin size equals 1, so the ordinate of the p.d.f. can be read as the relative probability of occurrence. (*d*) Variance in **d**, the distance between the best class and the mean, is translated into variance in the parameter of the distribution. This is achieved in two steps; first expressing **d** in units measured from the anchor of the distribution, and then scaling the variance to units appropriate for the position of the mean. $N \cong 10^3$; $s = 0{\cdot}005$; $\mu = 2$; $t + d = 30 + 18 = \mu/s - k$.

non-centralized p.d.f.s. A final translation back to the right will restore the corrected p.m.f.

Let **d** be a random variable that equals the difference between the best class and the mean. It is reasonable to assume that the distribution of **d** is Gaussian. Yet this must be only an approximation, for the normal distribution, besides having support on the entire real axis, imposes no restrictions on the relationship between the mean and the variance. Because for any set of initial conditions N, s and μ there exists exactly one equilibrium distribution, there must be an *a priori* relationship between the mean and variance of **d**. It is this parametric relationship between the mean and the variance that will allow the identification of a specific parameterized distribution for each combination N, s and μ. There is a rich literature to suggest (e.g. Johnson *et al.* 1993, ch. 17) that when there is evidence that the normal distribution is indeed only an

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

(asymptotically) approximating distribution for continuous non-negative variates, then the actual distribution may be better fit by a gamma distribution (Fig. 3*c*).

The two-parameter gamma distribution has a p.d.f.

$$p(x) = \frac{x^{\alpha-1}\, e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \qquad \alpha > 0, \beta > 0, x > 0, \tag{4}$$

with mean $\mathrm{E}(\mathrm{X}) = \alpha\beta$ and variance $\mathrm{Var}(\mathrm{X}) = \alpha\beta^2$ for shape parameter α and scale parameter β. Because **d** represents a difference, while the parameter λ is more appropriately measured from 0 to the mean, construct a new random variable **t**, that has the same distribution as **d**, but is now measured $b-k$ units from the origin (Fig. 3*d*). This maps the difference **d** into a random variable centered at the expected (shifted) position of the best class $t = \mathrm{E}(\mathbf{t}) = b-k$. Because **d** units from the mean is centered at the same position as **t** units from the origin, there is no change in the scale of units when transforming from **d** to **t**. To set the scale of the distribution of **t** so the ordinate can be read as p.m.f., let $\beta = t/(\mu/s-k)$. Because $\mathrm{E}(\mathbf{t}) = \alpha\beta = t$, this implies that $\alpha = \mu/s-k$.

Let λ be a random variable that stands for the parameter of the distribution of the number of mutations per individual. Then if the variance in λ is directly proportional to the variance in **d** (and consequently **t**), λ may be constructed by mapping $\mathbf{t} \to \lambda$ while maintaining the ratio of the variance to the mean. By setting $\mathrm{E}(\lambda) = \lambda = \mu/s-k$ and keeping β constant, λ is centered around its expectation while the variance to mean ratio is preserved. With $\beta = t/(\mu/s-k), \alpha = (\mu/s-k)^2/t$ and

$$\mathrm{E}(\lambda) = \mu/s-k,$$
$$\mathrm{Var}(\lambda) = t.$$

In a classic paper, Greenwood & Yule (1920) showed that when the parameter of Poisson distribution is distributed as a gamma variate, the resulting compound distribution is the negative binomial. In a notation consistent with (eqn 4), the p.m.f. is

$$\Pr(\mathrm{X} = x) = \binom{\alpha+x-1}{\alpha-1}\left(\frac{\beta}{1+\beta}\right)^{x}\left(\frac{1}{1+\beta}\right)^{\alpha},$$

with mean and variance

$$\mathrm{E}(\mathrm{X}) = \alpha\beta = \mu/s-k,$$
$$\mathrm{Var}(\mathrm{X}) = \alpha\beta+\alpha\beta^2 = \mu/s-k+t.$$

Translating k units to the right, the final distribution for the number of mutations per individual is a shifted negative binomial

$$\Pr(\mathrm{X} = x) = \begin{cases} \binom{\alpha+(x-k)-1}{\alpha-1}\left(\frac{\beta}{1+\beta}\right)^{(x-k)}\left(\frac{1}{1+\beta}\right)^{\alpha} & x \geqslant k \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

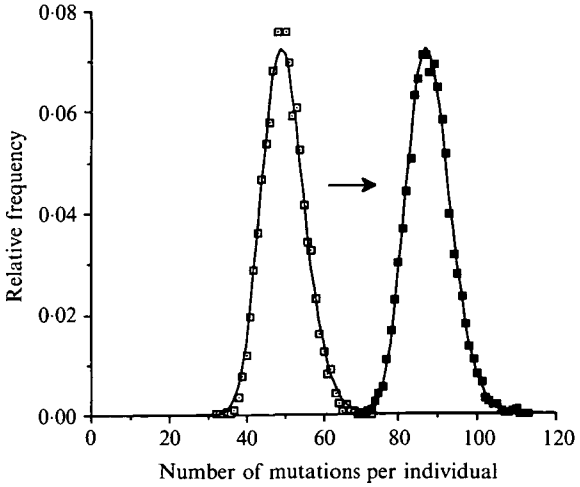

Fig. 4. Continuous-line representations of shifted negative binomials as they fit the simulation in Fig. 1. The negative binomial is derived from the Poisson by taking into account the variance in the distance between the best class and the mean. The location of the right-hand distribution uses the predicted rate of the ratchet (see text). Both p.m.f.s are parameterized solely from knowledge of N, s, and μ; neither was placed by prior reference to the simulations. $\mu/s-k = 23$; $k = 27$; $b = 35$; $t = 8$; $\alpha = 23^2/8$; $\beta = 8/23$.

with α and β defined as above, $\mathrm{E}(\mathrm{X}) = \mu/s$, and $\mathrm{Var}(\mathrm{X}) = \mu/s-k+t$.

The result that the variance of the mixed distribution is approximately equal to the (unshifted) mean plus the variance of the mixing distribution is consistent with other derivations that yield similar mixed distributions. For example, Kemp & Kemp (1965) and Kemp & Kemp (1966) found that if a mixing distribution on a Poisson variate is normal with mean λ and variance σ_λ^2, then, under some mild restrictions, the resulting mixed distribution is Hermite with mean λ and variance $\lambda+\sigma_\lambda^2$.

Figure 4 shows how (eqn 5) fits the simulations reported in Figs. 1 and 2. Table 1, which also includes predictions on the position of the worst class, shows how (eqn 5) fits other simulations. The worst class is predicted to be the last class in (eqn 5) that has an expectation of at least one individual.

5. The rate of the ratchet

Given that an approximate p.m.f. for the number of mutations per individual is a shifted negative binomial, one can formulate an expression for the rate of Muller's ratchet. The rate of the ratchet is approximately

$$R \cong \mu - s\sigma_n^2, \tag{6}$$

where σ_n^2 is the variance in the number of mutations per individual (eqn 14, Haigh, 1978; eqn 16, Pamilo *et al.* 1987). Haigh's analysis identifies $\sigma_n^2 = \theta = \mu/s$ as the criterion for equilibrium, with finite populations approaching this value very quickly as $N \to \infty$.

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

Equations (1), (2) and (in the limit) (eqn 5) reduce to Haigh's solution and predict $\sigma_n^2 = \theta$ when the zero class is present ($k = 0$). But whenever the zero class is missing while the mean still equals θ, the variance of the distribution falls below the Poisson expectation, and under these conditions, we had no reliable theory for estimating σ_n^2.

From (eqn 1), $\lambda = \mu/s - k$ is both the parameter of the distribution and the variance. Substituting this into (eqn 6) yields an upper bound estimate of the rate of the ratchet as

$$R = sk. \tag{7}$$

From (eqn 5), the correction for stochasticity in the distance between the best class and the mean yields

$$R = sk - st. \tag{8}$$

The addition of stochasticity in the distance between the best class and the mean increases the variance of the population proportional to the distance that the expected best class exceeds the first non-zero probability. This stochastically induced increase in the variance of the population decreases the rate of the ratchet. Table 1 compares the rate predicted by (eqn 8) with the observed rate.

One may also note that although the variance in the position of the best class makes an important contribution to determining the rate of the ratchet, the over-all importance of stochasticity in driving the ratchet diminishes as the mutation rate increases. As the mutation rate becomes large, the load increases, and the number of offspring ($N = N_A e^{\mu}$) needed from a pool of N_A adults grows exponentially. For a finite population, as μ becomes large the expected frequency of the zero class quickly approaches zero and the probability that $k > 0$ approaches unity, guaranteeing a non-stochastic component to the ratchet. Concurrently, the probability that members of the best class acquire no new mutations becomes vanishingly small ($e^{-\mu}$). Under a large mutation rate not only is the strength of selection on the best class too weak, but in each generation, few individuals in the best class will escape acquiring new mutations. In this situation, it is mutation pressure that primarily diminishes the best class and drives the ratchet.

6. Rate of decay in mean fitness

Given the rate of mutation accumulation, one can extend it to the rate of decay in mean fitness. Temporarily ignoring drift and assuming constant selection coefficients (an assumption examined in the next section), the mean population fitness for a large finite population is approximately its equilibrium expectation times the effect of accumulated mutations, $e^{-\mu}(1-s)^{R\tau}$, where τ is measured in time since the population first reaches its mean fitness of $e^{-\mu}$. $R\tau$ equals the number of turns of the ratchet. The regression of $e^{-\mu}(1-s)^{R\tau}$ on $R\tau$ defines mean population fitness as a function of the number of accumulated mutations, and the rate the population traverses this curve, i.e. a regression on τ, is a function of N, s and μ. Because fitness in this model is exclusively multiplicative, the rate of decay is linear on a log scale, and thus

$$\bar{w}(\tau) = e^{-\mu}(1-s)^{R\tau},$$

$$\ln(\bar{w}(\tau)) = -\mu + R\tau \ln(1-s)$$

and

$$\frac{d\ln(\bar{w}(\tau))}{d\tau} = R\ln(1-s) \quad \text{for } R \text{ independent of } \tau. \tag{9}$$

7. Variable selection coefficients

The logic of the preceding analysis can be applied to the more realistic case where mutations assume a distribution of deleterious effects (Fig. 5). In a population where incoming selection coefficients are distributed as negative exponential variates (Ohta, 1977; Gillespie, 1991), the existence of mutations of small effect makes it even more likely that every individual will have at least one mutation. This increases the likelihood of an absent zero class and consequently a hypodispersed population.

For constant selection coefficients, the variance in fitness amongst individuals is determined entirely by the variance in the number of mutations. Variable selection coefficients change this relationship and introduce three new aspects to mutation accumulation. First, because the best class (the class with the fewest mutations), is not necessarily the most fit class, selection on the most fit individuals does not minimize mutation accumulation *per se*: mutation accumulation

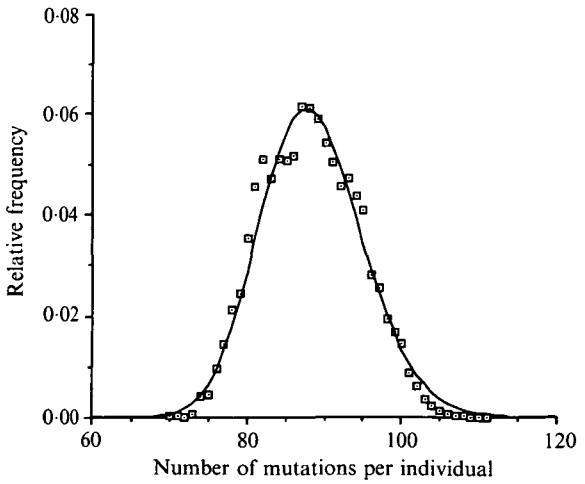

Fig. 5. The same parameter values as used in Fig. 1 except with variable selection coefficients. $\bar{s}_{seg} = 0{\cdot}0113$; $\theta = 88{\cdot}50$; $k = 58$; $b = 71$.

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

Table 2. *Similar to Table 1, except incoming mutations are distributed as approximate negative exponential variates.* $\bar{s}_{inc}$ *is the mean incoming selection coefficient;* $\bar{s}_{seg}$ *is the mean observed selection coefficient amongst segregating loci over the last* N/2 *generations*

				Best class		Worst class	
N	$\bar{s}_{inc}$	μ	k	Predicted	Observed	Predicted	Observed
500	0·005	0·5	99	112	112·61	142	141·09
		1	198	218	216·40	252	251·14
		2	401	429	428·42	467	465·66
	0·01	0·5	48	57	56·86	83	82·74
		1	99	112	111·61	142	142·28
		2	197	217	216·31	251	249·39
	0·02	0·5	23	29	29·29	51	50·36
		1	49	58	58·00	84	84·10
		2	98	111	110·86	140	139·40
1000	0·005	0·5	99	114	113·90	148	148·09
		1	203	225	223·58	264	265·23
		2	398	430	428·81	475	474·87
	0·01	0·5	48	57	57·02	87	87·56
		1	101	116	115·71	151	151·73
		2	201	223	222·16	262	262·02
	0·02	0·5	24	30	30·33	56	55·88
		1	50	59	58·57	89	90·00
		2	99	114	113·46	148	148·27

		Variance			Rate of the ratchet	
		Poisson	Predicted	Observed	Predicted	Observed
$\bar{s}_{seg} \times 10^3$	$t = b - k$	$\theta = \mu/\bar{s}_{seg}$	$\mu/\bar{s}_{seg} - k + t$	$\sigma_n^2 \pm$ S.E.	$R = \bar{s}_{seg} k - \bar{s}_{seg} t$	$R_{obs} \pm$ S.E. $\times 10^2$
4·002	13	124·94	38·94	34·33 ± 10·00	0·34	0·36 ± 3·01
4·312	20	231·91	53·91	54·24 ± 9·44	0·77	0·80 ± 6·96
4·496	28	444·84	71·84	70·61 ± 17·80	1·68	1·71 ± 11·66
7·399	9	67·58	28·58	28·89 ± 12·68	0·29	0·32 ± 3·44
8·014	13	124·78	38·78	39·74 ± 8·61	0·69	0·73 ± 4·27
8·664	20	230·84	53·84	52·62 ± 9·63	1·53	1·64 ± ·4·54
13·17	6	37·97	20·97	16·35 ± 2·00	0·22	0·28 ± 2·04
14·51	9	68·92	28·92	27·00 ± 4·50	0·58	0·63 ± 3·33
16·22	13	123·31	38·31	35·24 ± 6·92	1·38	1·46 ± 6·98
3·899	15	128·24	44·24	42·73 ± 7·20	0·33	0·35 ± 2·57
4·148	22	241·08	60·08	65·11 ± 8·40	0·75	0·76 ± 2·91
4·456	32	448·83	82·83	95·29 ± 14·87	1·63	1·65 ± 3·84
7·161	9	69·82	30·82	33·98 ± 4·60	0·28	0·30 ± 2·74
7·652	15	130·69	44·69	46·34 ± 11·10	0·66	0·68 ± 3·24
8·357	22	239·32	60·32	61·48 ± 9·09	1·50	1·54 ± 4·45
12·26	6	40·78	22·78	21·44 ± 3·88	0·22	0·25 ± 2·48
13·93	9	71·79	30·79	33·99 ± 4·87	0·57	0·59 ± 2·77
15·59	15	128·29	44·29	44·69 ± 7·15	1·31	1·42 ± 3·05

is minimized as a correlated response. Secondly, the introduction of mutations of differing effect at differing rates results in a lower mean selection coefficient amongst segregating alleles (e.g. see Keightley, 1994). This means that constant s models underpredict the mean number of mutations per individual. Thirdly, the introduction of mutations of effect $s \ll N_e^{-1}$ (Kimura, 1968, 1983, p. 44; Li, 1978) will result in a class of alleles that can drift out of mutation–selection balance, thereby augmenting mutation accumulation.

With variable selection coefficients, the fact that the incoming distribution of mutations does not reflect the equilibrium distribution introduces two additional summary statistics that affect the rate of the ratchet and the rate of decline in mean fitness: $\bar{s}_{seg}$, the mean selection coefficient of all alleles that are currently segregating, and $\bar{s}_{fixed}$, the mean selection coefficient of all alleles fixed in the population. The latter is approximated by the mean selection coefficient of all fixed loci when measured after a sufficiently long period of time.

(i) *Rate of the ratchet*

If most alleles are being held in mutation–selection balance, then a mean fitness of $e^{-\mu} \cong (1 - \bar{s}_{seg})^n$ implies

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

$\bar{n} \cong \mu/\bar{s}_{seg} = \theta$ for any $\bar{s}_{seg} \ll 1$, where $\bar{n}$ is the average number of mutations per individual. Thus for variable selection coefficients, the expected mean number of mutations per individual is determined by $\bar{s}_{seg}$, the average effect of all mutations segregating in the population, and θ is redefined as

$$\theta \cong \mu/\bar{s}_{seg}. \tag{10}$$

The same approximation can be derived as a limiting case for a diffusion model following the form of Kimura (1969, 1983, p. 239) and Ewens (1979). Even with mutations drifting out of mutation–selection balance (eqn 10) will still estimate the mean number of mutations per individual, though the algebraic relationship between selection coefficients assigned to alleles and a calculated fitness statistic for the population will no longer be indicative of the actual strength of selection. Drift will invalidate the expected distance from the best class to the mean as a measure of the ability of the population to resist mutation accumulation, and for this reason (eqn 8) using $s = \bar{s}_{seg}$ will tend to underestimate the rate of the ratchet. This will be true for both constant and variable s models whenever s or $\bar{s}_{seg}$ are implicitly assumed to reflect the true strength of selection.

To circumvent this, $\bar{s}_{seg}$ should be replaced by a statistic that accurately estimates the true strength of selection, not just the arithmetic average selection coefficient of segregating alleles. But an analytical prediction for $\bar{s}_{seg}$ is complicated by simultaneously occuring background selection affecting the long-term net effect of selection and the mean time to fixation (Hill & Roberston, 1966; Charlesworth, B. *et al.* 1993; Charlesworth, 1994; Barton, 1994). Despite this, the general inequality $\bar{s}_{inc} \geqslant \bar{s}_{seg} \geqslant \bar{s}_{fixed}$ (where $\bar{s}_{inc}$ is the mean incoming selection coefficient) allows a qualitative prediction of how constant s models compare to variable s models.

Because the equality is met only in the degenerative case (i.e. when s is constant or when all mutations can freely drift), the inequality $\bar{s}_{inc} > \bar{s}_{seg}$ increases the likelihood that a variable s population will have a missing zero class. This reflects the verbal argument that to the degree that mutations of small effect can enter a population, it is increasingly unlikely that there will be any mutation-free individuals. As $\bar{s}_{seg}$ decreases linearly, the inequality $N e^{-\theta} < 1$ is satisfied for an exponentially larger range of N. This causes k to increase faster than $\bar{s}_{seg}$ decreases. This results in a faster ratchet for populations with variable selection coefficients.

When only a few individuals are expected in the best class, the time to lose the best class stochastically can still be small compared to the non-stochastic component. This is demonstrated in Table 2 where the rate of the ratchet using the observed $\bar{s}_{seg}$ and (eqn 10) are still reasonably approximated by a shifted negative binomial.

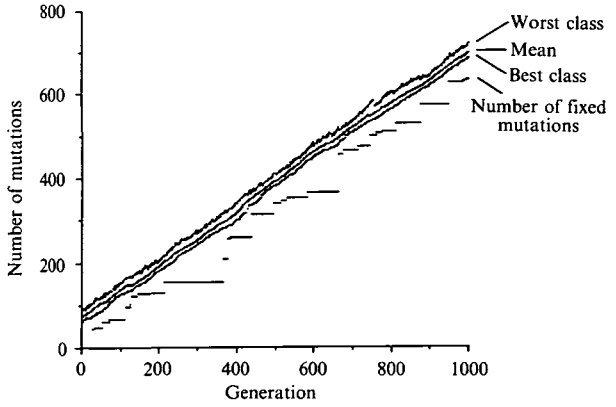

Fig. 6. A sample variable s run showing the increase in the number of fixed mutations, and the minimum, mean and maximum number of mutations per individual over time. The saltatory increase in the number of fixed mutations is indicative of how selective sweeps fix multiple mutations simultaneously in asexual populations. The observed y-intercept and slope (the rate of the ratchet) is $y = 71{\cdot}71 + 0{\cdot}63\tau (r^2 = 0{\cdot}992)$. The predicted regression using the observed $\bar{s}_{seg}$ is $y = \mu/\bar{s}_{seg} + R\tau = 72{\cdot}44 + 0{\cdot}55\tau$. $N \cong 10^3$; $\bar{s}_{inc} = 0{\cdot}02$; $\mu = 1$; $\bar{s}_{seg} = 0{\cdot}01381$.

(ii) *Rate of decay in log mean fitness*

Mutations of very small effect accumulate almost freely, yet individually add little to the decay in mean fitness. For alleles at strictly neutral loci, their influx drives mutation accumulation, yet their neutrality causes no decay in mean fitness. At the other extreme, mutations of very large effect experience an almost deterministic efficiency of selection, so their contribution to the decline in mean fitness is also small. In both cases, population mean fitness declines only very slowly beyond the infinite-size mutation–selection equilibrium. Between these two extremes are values of s small enough to accrue, yet large enough to substantially effect the rate of decay in mean fitness (Kimura *et al.* 1963; Charlesworth, D. *et al.* 1993; Gabriel, *et al.* 1993). The prediction then, is that there exists a critical incoming value of $\bar{s}_{inc}$ that maximizes the magnitude of the rate of decay in mean fitness.

In the derivation for the rate of change in log mean fitness (eqn 9), the determining factor is the average effect of accumulated alleles. Without back mutation pressure, selective sweeps will equilibrate the rate of fixation with the rate of the ratchet (Fig. 6), and therefore

$$\frac{\mathrm{d} \ln (\bar{w}(\tau))}{\mathrm{d}\tau} = R \ln (1 - \bar{s}_{fixed}).$$

For $\bar{s}_{inc} \ll N_e^{-1}$, both constant and variable s populations will decay at approximately the same rate. As $\bar{s}_{inc}$ increases, variable s populations benefit from a lower $\bar{s}_{fixed}$, and therefore decay at a slower rate. But as $\bar{s}_{inc}$ increases further still, variable s populations continue to have some frequency of mutations of small effect that can enter and fix, while constant s populations approach a deterministic

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

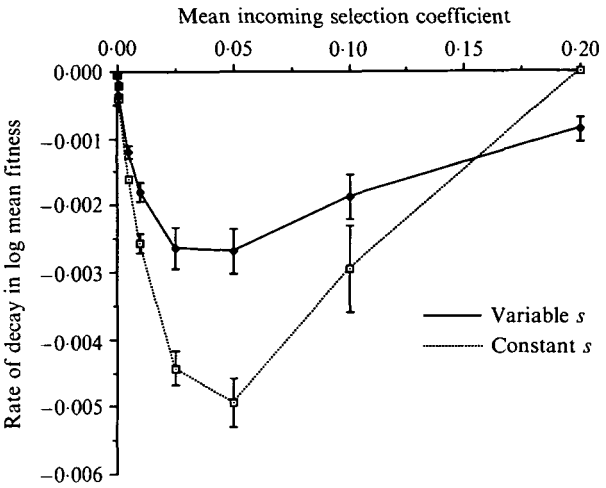

Fig. 7. Rate of decay in log mean fitness as a function of the mean incoming selection coefficient. Each data point is the mean of ten independent runs with its standard error. $N \cong 10^3$; $\bar{s}_{inc} = 0{\cdot}02$; $\mu = 0{\cdot}5$.

efficiency in selection. At some critical value of $\bar{s}_{inc}$ these trade-offs balance, and there are equivalent rates of decay (Fig. 7). As expected though, for any given incoming selection coefficient, an increase in the population size or a decrease in the mutation rate always results in a slower rate of decay in mean fitness.

8. Discussion

(i) *Caveats of the theory*

The distribution in (eqn 5) relies on the accurate identification of the position of the best class. If θ is small, then the expected size of each class changes rapidly for classes b, $b+1$, $b+2$, etc. and the time-average size of the best class may not be well approximated by b. For example, for $N = 500$, $s = 0{\cdot}01$ and $\mu = 0{\cdot}1$, the expected best class is $b = 3$, while the observed best class is $b_{obs} = 5$. This is also true if N is very small, though it is doubtful that in either case any increase in resolution above (eqn 1) and $R = sk$ is justified.

The accurate prediction of the position of the best class leads to the use of a truncated Poisson in (eqn 2), though this is not incorporated in (eqn 5). This reflects the fact that (eqn 5) (and most parametric distributions) are more sensitive to methods that lead to their parameterization, than to their own truncation. Small changes in the form of the original distribution, e.g. as a Poisson or a truncated Poisson, or in the mixing distribution, be it a gamma, a truncated gamma or a truncated normal, will yield slightly different mixed distributions as a result. There is a wide range of overlap between the negative binomial, the hyper-Poisson distribution (Bardwell & Crow, 1964; Crow & Bardwell, 1965), the Hermite (McKendrick, 1926), and others as approximating distributions. All these distributions are themselves special cases of more general distributions. The use of a specific mixing distribution is partly justified by the emphasis placed on assumptions leading to its derivation, and partly by the parsimony of the resulting mixed distribution to explain the data. See for example, Barton's (1966) pointed review of Staff (1964), where he strongly advocates the use of a negative binomial in the absence of compelling reasons otherwise.

(ii) *Implications of the theory*

Haigh's (1978) analysis showed that if $N e^{-\theta} \lesssim 25$, then otherwise stable asexual populations are subject to mutation accumulation at a rate proportional to the rate of the stochastic loss of the best class. This analysis shows that if $N e^{-\theta} < 1$, then the equilibrium distribution of mutations necessary for mutation–selection balance is never attained and mutations accumulate due to both stochastic and non-stochastic pressures.

This leads to a model for the distribution of the number of mutations per individual. Under the caveats described above, the distribution is expected to be Poisson, albeit a shifted Poisson of parameter $\lambda = \mu/s - k$. After adjusting for variance in the parameter of the distribution (a phenomenon introduced by the Monte Carlo modeling process), the distribution becomes relatively hyperdispersed, and becomes a shifted negative binomial.

There are three points that revise our thinking about asexual populations with multiplicative fitness. First, in finite, asexual populations where we do not expect at least one 'mutation-free' individual relative to a mutation-free standard, mutation–selection balance will not be achieved and there is no stationary distribution of mutations. This is due specifically to the strength of selection on the best class being too weak to counter mutation pressure. The effect can be large: for example, using the previous estimates of $\mu = 1$ and $s = 0{\cdot}02$, for a population of 10^9 individuals ($\theta = \mu/s = 50$, $k = 14$, $b = 21$), the ratio of the predicted variance to the necessary variance needed to halt mutation accumulation is

$$\frac{\sigma_n^2}{\theta} \cong \frac{\mu/s - k + b - k}{\mu/s} = \frac{43}{50} = 0{\cdot}86.$$

Mutations will accumulate not just because the best individuals may be lost by chance, but because they are destined to be lost due to their insufficient selective advantage relative to the other classes. Muller's ratchet is not reserved for small populations, but will happen in all populations that cannot maintain mutation-free individuals.

Secondly, this consideration provides an estimate of the rate Muller's ratchet ($R = sk - st$) when $N e^{-\theta} < 1$. As population size increases, the absolute rate of the ratchet decreases, but ironically stochasticity becomes

https://doi.org/10.1017/S0016672300034686 Published online by Cambridge University Press

increasingly important in determining this rate. Stated alternatively, if $1 < N e^{-\theta} < 25$, then stochasticity primarily drives Muller's ratchet (Haigh, 1978). As population size decreases below the critical inequality $N e^{-\theta} < 1$, it is mutation pressure that increasingly drives the ratchet.

The rate of the ratchet is mainly determined by the ratio of the influx to efflux of mutations ($\theta = \mu/s$). For small values of μ (e.g. $\mu \cong 0{\cdot}0033$ for DNA microbes [Drake 1991]) and $s = 0{\cdot}02$, the ratchet is not expected to operate, since a large number of individuals will be in the zero class ($N e^{-\theta} > 1$ implies $N > 1{\cdot}18$). For $\mu = 0{\cdot}1$, populations could maintain a zero class with only 150 individuals. But for $\mu \cong 0{\cdot}4$ (based on human data, Koeberl *et al.* 1990; Sommer, 1992; though see Kondrashov & Crow, 1993), populations would already need an effective size of 5×10^8 to achieve mutation–selection balance.

The preceding is conservative in terms of the rate of mutation accumulation, for by using constant selection coefficients it underestimates the rate to the ratchet. Correcting for variable selection coefficients, even if the average segregating mutation has a selection coefficient as high as 60% of the incoming value ($\bar{s}_{\text{seg}} = 0{\cdot}01226$, $N = 10^3$, Table 2), the necessary number of individuals for $\mu = 0{\cdot}4$ jumps to $1{\cdot}5 \times 10^{14}$. And with this increase in N, the concordant increase in the efficiency of selection will decrease $\bar{s}_{\text{seg}}$ further still. For values of μ close to unity (Charlesworth *et al.* 1990; Crow, 1993*a*, *b*), no population, with either constant or variable selection coefficients, would ever be large enough to be able to maintain asexual reproduction and still achieve the necessary strength of selection to counter mutation pressure ($s = 0{\cdot}02 \Rightarrow N > 5 \times 10^{21}$). For organisms with large genomes, a fast ratchet may be an inescapable consequence of asexual reproduction.

The simplification that Muller's ratchet is due to the stochastic loss of the best class in small populations may miss the primary importance of the ratchet in the evolution of metazoans and metaphyta. In DNA microbes, increased mutation rates (correlated with increasingly larger genomes) may be compensated by more complex error-correction mechanisms (Drake, 1991). Yet as the number of genes increases further still, recombination becomes a more plausible load reducing mechanism (Felsenstein, 1974 and references therein; Kondrashov, 1982). Haigh's (1978) identification of $N e^{-\theta}$ as a pivotal quantity affecting the rate of the ratchet sets an important quantitative bound on the parameter-space where recombination is expected to evolve. Quantification of this rate can help to determine the relative strength of selection for a recombination modifier; i.e. larger values of k reflect a paucity of variance and should be correlated with stronger selection for the evolution of recombination.

The restriction of Muller's ratchet from being operational in very large populations has been used as evidence that it may not be a sufficient reason to drive the evolution of sex (Crow, 1994). An interpretation of the results presented here shows that if genome size increases faster than error–correction mechanisms evolve, then even large populations could generate strong selection for the evolution of load reducing mechanisms.

Thirdly, variable selection coefficients result in a faster ratchet, but for small values of s, a slower decay in mean fitness. With variable selection coefficients it is more likely that asexual populations will not be able to achieve mutation–selection balance, but this comes at a smaller than expected cost. Because of this study's omission of epistasis, beneficial mutations, compensatory mutations, fluctuating selection coefficients, etc., direct applications to natural asexual populations are cautionary. But to the degree that unconditionally deleterious-allele models of this type are appropriate, it implies that populations of asexual organisms with large genomes are acquiring mutations faster than previously expected, for reasons other than previously expected, yet should able to persist longer than previously expected.

I thank William Rice for bringing this problem to my attention, and for originally pointing out the importance of a reduction in the steady-state strength of selection on the best class in driving the ratchet. I also thank Alan Hastings, Austin Burt, Michael Lynch and two additional referees for their comments on earlier revisions that greatly improved the quality of this manuscript.

References

Bardwell, G. E. & Crow, E. L. (1964). A two-parameter family of hyper-Poisson distributions. *Journal of the American Statistical Association* **59**, 133–141.

Barton, D. E. (1966). *Review of* P. J. Staff: The displaced Poisson distribution. *Mathematical Reviews* **32**, 6583.

Barton, N. H. (1994). The reduction in fixation probability caused by substitutions at linked loci. *Genetical Research* **64**, 199–208.

Bell, G. (1982). *The Masterpiece of Nature*. Berkeley: University of California Press.

Charlesworth, B. (1990). Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* **55**, 199–221.

Charlesworth, B. (1994). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research* **63**, 213–227.

Charlesworth, B., Charlesworth, D. & Morgan, M. T. (1990). Genetic loads and estimates of mutation rates in highly inbred plant populations. *Nature* **347**, 380–382.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*. **134**, 1289–1303.

Charlesworth, D., Morgan, M. T. & Charlesworth, B. (1993). Mutation accumulation in finite outbreeding and inbreeding populations. *Genetical Research* **61**, 39–56.

Crow, E. L. & Bardwell, G. E. (1965). Estimation of parameters of the hyper-Poisson distributions. In: *Classical and Contagious Discrete Distributions* (ed. Patil, G. P.), pp. 127–140. Statistical Publishing Society. Calcutta. Oxford: Pergamon Press.

Crow, J. F. (1993*a*). How much do we know about spontaneous human mutation rates? *Environmental and Molecular Mutagenesis* **21**, 122–129.

Crow, J. F. (1993*b*). How much do we know about spontaneous human mutation rates? *Correction. Environmental and Molecular Mutagenesis* **21**, 389.
Crow, J. F. (1994). Advantages of sexual reproduction. *Developmental Genetics* **15**, 205–213.
Crow, J. F. & Simmons, M. J. (1983). The mutation load in Drosophila. In: *The Genetics and Biology of* Drosophila. Ashburner, M. *et al.* (eds.), pp. 1–35. London: Academic Press.
Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Science USA* **88**, 7160–7164.
Ewens, W. J. (1979). *Mathematical Population Genetics*, p. 239. New York: Springer-Verlag.
Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* **78**, 737–756.
Gabriel, W., Lynch, M. & Bürger, R. (1993). Muller's ratchet and mutational meltdowns. *Evolution* **47**, 1744–1757.
Gillespie, J. H. (1991). *The Causes of Molecular Evolution*, pp. 262–266. New York: Oxford University Press.
Greenwood, M. & Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Society, Series* A. **83**, 255–279.
Haigh, J. (1978). The accumulation of deleterious genes in a population – Muller's ratchet. *Theoretical Population Biology* **14**, 251–267.
Haight, F. A. (1967). *Handbook of the Poisson Distribution*. New York: John Wiley & Sons.
Haldane, J. B. S. (1937). The effect of variation on fitness. *American Naturalist* **71**, 337–349.
Higgs, P. G. (1994). Error thresholds and stationary mutant distributions in multi-locus diploid genetics models. *Genetical Research* **63**, 63–78.
Hill, W. G. & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research* **8**, 269–294.
Johnson, N. L., Kotz, S. & Kemp, A. W. (1993). *Univariate Discrete Distributions*. 2nd ed. New York: John Wiley & Sons.
Keightley, P. D. (1994). The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**, 1315–1322.
Kemp, A. W. & Kemp, C. D. (1966). An alternative derivation of the Hermite distribution. *Biometrika* **53**, 627–628.
Kemp, C. D. & Kemp, A. W. (1965). Some properties of the 'Hermite' distribution. *Biometrika* **52**, 381–394.
Kimura, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research* **11**, 247–269.
Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to the steady flux of mutations. *Genetics* **61**, 893–903.
Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
Kimura, M. & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351.
Kimura, M., Maruyama, T. & Crow, J. F. (1963). The mutation load in small populations. *Genetics* **48**, 1303–1312.
Koeberl, D. D., Bottema, C. D. K., Ketterling, R. P., Bridge, P. J., Lillicrap, D. P. & Sommer, S. S. (1990). Mutations causing Hemophila B: Direct estimate of the underlying rates of spontaneous germ-line transitions, transversions, and deletions in a human gene. *American Journal of Human Genetics* **47**, 202–217.
Kondrashov, A. S. (1982). Selection against harmful mutations in large sexual and asexual populations. *Genetical Research* **40**, 325–332.
Kondrashov, A. S. & Crow, J. F. (1993). A molecular approach to estimating the human deleterious mutation rate. *Human Mutation* **2**, 229–234.
Li, W.-H. (1978). Maintenance of genetic variability under the joint effect of mutation, selection, and random drift. *Genetics* **90**, 349–82.
Lynch, M. & Gabriel, W. (1990). Mutation load and the survival of small populations. *Evolution* **44**, 1725–1737.
Maynard Smith, J. (1978). *The Evolution of Sex*, pp. 33–36. Cambridge: Cambridge University Press.
McKendrick, A. G. (1926). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society* **44**, 98–130.
Muller, H. (1950). Our load of mutations. *American Journal of Human Genetics* **2**, 111–176.
Muller, H. (1964). The relation of recombination to mutational advance. *Mutational Research* **1**, 2–9.
Ohta, T. (1977). Extension to the neutral mutation random drift hypothesis. In: *Molecular Evolution and Polymorphism* (ed. M. Kimura), pp. 148–167. National Institutes of Genetics.
Pamilo, P., Nei, M. & Li, W.-H. (1987). Accumulation of mutations in sexual and asexual populations. *Genetical Research* **49**, 135–146.
Simmons, M. J. & Crow, J. F. (1977). Mutations affecting fitness in *Drosophila* populations. *Annual Review of Genetics* **11**, 49–78.
Sommer, S. S. (1992). Assessing the underlying pattern of human germline mutations: lessons from the factor IX gene. *The FASEB Journal* **6**, 2767–2774.
Staff, P. J. (1964). The displaced Poisson distribution. *Australian Journal of Statistics* **6**, 12–20.
Staff, P. J. (1967). The displaced Poisson distribution – Region B. *Journal of the American Statistical Association* **62**, 643–654.
Stephan, W., Chao, L. & Smale, J. G. (1993). The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genetical Research* **61**, 225–231.
Teicher, H. (1960). On the mixture of distributions. *Annals of Mathematical Statistics* **31**, 55–73.