# The Generalized Haldane (GH) model tracking population size changes and resolving paradoxes of genetic drift

Yongsen Ruan[*,1], Xiaopei Wang[1], Mei Hou[1], Liying Huang[1], Wenjie Diao[1], Miles Tracy[1], Shuhua Xu[2], Weiwei Zhai[3], Zhongqi Liufu[1], Haijun Wen[*,1] and Chung-I Wu[*,1]

**Affiliations**

[1]State Key Laboratory of Biocontrol, Innovation Center for Evolutionary Synthetic Biology, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China.

[2]Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai 200438, China.

[3]Institute of Zoology, Chinese Academy of Sciences, Beijing, China

*Corresponding authors: ruanys3@mail.sysu.edu.cn, wenhj5@mail.sysu.edu.cn, ciwu@uchicago.edu.

**Abstract**

Population genetic models, such as the Wright-Fisher (WF) model, track relative gene frequencies. The absolute gene copy number, or population size ($N$), is supplied externally for tracking genetic drift. JBS Haldane (1927) proposed an alternative model based on the branching process, whereby each gene copy is transmitted to $K$ descendants with the mean and variance of $E(K)$ and $V(K)$. In this model, $E(K)$ governs $N$, while $V(K)/N$ governs genetic drift. Nevertheless, as the branching process allows $N$ to drift unboundedly, a Generalized Haldane (GH) model that regulates N more tightly is proposed. The GH model can account for several paradoxes of molecular evolution. Notably, genetic drift may often become stronger as $N$ becomes larger in the ecological setting, thus contradicting the general view. In particular, a very small population growing exponentially experiences little drift. Interestingly, when the population grows and $N$ oscillates near the carrying capacity, the paradoxical trend is also observed in both field works and laboratory experiments. This paradox whereby population size in genetics ($N_e$) and ecology ($N$) could be negatively correlated is resolved by the GH model. Additional paradoxes include ii) The two sexes experiencing drift differently; iii) Genetic drift of advantageous mutations being independent of $N$; iv) Multi-copy gene systems (viruses, mitochondria, etc.) having no definable $N_e$ (for effective $N$). In brief, the GH model defines genetic drift simply as $V(K)$, or $V(K)/N$ averaged over the population. It represents an attempt at integrating genetical and ecological analyses into one framework.

## Introduction

Genetic drift is defined as the random changes in frequency of genetic variants (Crow and

Kimura 1970; Hartl and Clark 1997; Lynch, et al. 2016). Much, or even most, of DNA sequence evolution is governed by the random drift of neutral mutations. Nevertheless, even non-neutral variants are subjected to genetic drift as advantageous mutations often have a small probability of being fixed. If the strength of genetic drift is under-estimated, random changes that are missed in the analysis would result in the over-estimation of other evolutionary forces of greater biological interest, including selection, mutation, migration, meiotic drive and so on. In particular, the conclusion of pervasive selection at the molecular level has been suspected to be due to the failure to fully account for genetic drift (Crow and Kimura 1970; Fu 1997; Li 1997; Charlesworth 2009; Lynch, et al. 2016; Chen, Yang, et al. 2022).

Hagedoorn and Hagedoorn (1921) may be the first to suggest that random forces (accidents, for example) could impact long-term evolution. In the same decade, models of genetic drift were developed along two lines. The one approach that became widely adopted is the Wright-Fisher (WF) model (Crow and Kimura 1970; Hartl and Clark 1997), formalized by Fisher (1930) and Wright (1931). They define genetic drift to be due to the random sampling of genes. Thus, in generation transition, the variance in the frequency of a variant allele, $x$, would be

$$V(x) = \frac{x(1-x)}{f(N)} = \frac{x(1-x)}{N_e} \quad (1)$$

where $N$ is the population size. (Here, we present the haploid model while a factor of 2 should be added to the diploid models.) The WF model has been modified in many directions (Wright 1969; Gillespie 1975; Eldon and Wakeley 2006; Chen, et al. 2017; Sackman, et al. 2019) whereby the effective population size, $N_e$, is used to accommodate various deviations from the ideal WF population. Note that $N$, and hence $N_e$, are imposed externally on the model as in the later Moran model (2008).

At about the same time, JBS Haldane introduced the branching process to model genetic drift (Haldane 1927; Haldane 1932). In the Haldane model, each gene copy is independently transmitted to $K$ descendants with the mean and variance of $E(K)$ and $V(K)$. Again, we present the haploid model, noting some added complexities with diploidy.

The branching process of the Haldane model can generate $N$ internally by formulating $N$ of the next generation as $E(N') = N \times E(K)$. Since Haldane's model (1927) resets $N'$ to a constant $N$, a Generalized Haldane (GH) model is proposed here that permits varying N under ecological regulation. Indeed, $E(K)$ is often negatively correlated with $N$ in actual populations (Smith and Slatkin 1973; Sibly and Hone 2002). The GH model thus tracks gene frequency changes as well as regulates $N$. In the Haldane model, $E(K)$ would determine the trajectory of $N$ changes and $V(K)$ would determine the strength of genetic drift. Thus, there is no drift if $V(K) = 0$ irrespective of N.

The GH model is then applied to several paradoxes of genetic drift (see the overview below). It is possible that mathematical modifications of the conventional models (i.e., WF and Moran models) may resolve these paradoxes as well. Nevertheless, the key question is whether these modifications, often highly sophisticated, are biologically feasible (see Discussion). In comparison, the Haldane model is intuitively appealing and deserves to be considered as an alternative approach to genetic drift. As we study the ever more complex biological systems in the genomic era, such an alternative would seem desirable.

## Results

### *Overview of the paradoxes of genetic drift*

Among the paradoxes, a most curious one is genetic drift in relation to changing population size. The WF model dictates stronger drift when $N$ is smaller. However, when $N$ is very small (say, $N = 10$ in a bacteria or yeast population) and increases to 20, 40, 80 and so on, there is in fact little drift in this exponential phase. Drift will intensify when $N$ grows to near the carrying capacity. This trend is the exact opposite of the standard view. The second paradox concerns the genetic drift of sex-linked genes in relation to sex-dependent breeding successes (Charlesworth and Charlesworth 2000; Bachtrog 2013; Cortez, et al. 2014; Wilson Sayres, et al. 2014; Makova, et al. 2024). A third paradox is about drift strength when selection is also at work. A new mutation with a selective advantage would always be fixed if there is no random drift. With drift, fixation is probabilistic, but the probability does not depend on $N$ or $N_e$. This curious property echoes the view that $N$ is a scaling factor of drift that is determined by $V(K)$ (see Results).

The power of the Haldane model in resolving these paradoxes is the focus of this study. Furthermore, the companion study (Wang et al. 2024) addresses a fourth paradox – the evolution of multi-copy gene systems, whereby evolution proceeds in two stages - between as well as within individuals. Multi-copy gene systems including viruses, mitochondria, transposons and ribosomal genes do not have easily definable Ne. Broadly speaking, even diploidy is also multi-copy systems (Silver 1985; Wu, et al. 1988; Wu, et al. 1989; Lindholm, et al. 2016; Courret, et al. 2019).

### *On the Haldane model of genetic drift*

In the original Haldane model, genetic drift is defined by the branching process (see Eq. (4) of Chen, et al. (2017)). Let $K$ be the number of progenies receiving the gene copy of interest from a parent; $K$ would follow the same distribution for all neutral variants. In haploids, $K$ is equivalent to the progeny number of each parent. In diploids, $K$ is the transmission success of each gene copy, which is separately tracked. By the Taylor approximations of the ratio of two variables (Kendall, et al. 2006), we obtain the approximation as

$$V(x) \sim \frac{x(1-x)}{N} \frac{V(K)}{E^2(K)} \qquad (2)$$

where $x$ is the variant frequency (see Supplementary Information). This is the same approximation, when $V(K)$ is not very large, as obtained by the WF model (Kimura and Crow 1964; see Chen et al. 2017 for details). For Eq. (2), we provide further extensive simulations on the accuracy in both the fixation probability and fixation time of neutral variants (see Supplementary Information). Note that N, constant or variable, is supplied externally to the WF model but it is tracked internally in the Haldane model (see the next section). The apparent equivalence of Eq. (2) makes a simple point that, between the WF and Haldane models, the mathematical results are often shared but the biological meanings may not be identical. (Similarly, the diffusion and coalescence theories are also distinct approaches to the same population genetic phenomena.)

In the Haldane model, there would be no genetic drift if $V(K) = 0$, regardless of the value of $N$. When $E(K) = 1$ and $N$ stays constant, we obtain

$$N_e = \frac{NE^2(K)}{V(K)} = \frac{N}{V(K)} \qquad (3)$$

Eq. (2) and Eq. (3) have been obtained for the WF model (Kimura and Crow 1963; Crow and Denniston 1988; Chen, et al. 2017). The issue will be highly significant when sex-dependent drift is analyzed.

**Table 1 here**

In **Table 1**, the data of progeny number from the literature indeed show over-dispersion in all taxa surveyed, i.e., $V(K) > E(K)$. In fact, $V(K) > 10\ E(K)$ can often be found, e.g., the ratio of $V(K)/E(K)$ among males of the mandrill (*Mandrillus sphinx*), the great reed warbler (*Acrocephalus arundinaceus*), and the rhesus macaque (*Macaca mulatta*) at 19.5, 12.6 and 11.3, respectively (Hasselquist 1995; Setchell, et al. 2005; Dubuc, et al. 2014). We have also found that $V(K)$ may be far larger than $E(K)$ in other biological systems such as viruses (Ruan, Luo, et al. 2021; Ruan, Wen, et al. 2021; Hou, et al. 2023).

## I. The first paradox – The paradox of changing *N*

### *1. Empirical demonstration and simulation*

This "paradox of changing $N$" is that "genetic drift increases when $N$ increases" in direct contradiction with Eq. (2). **Fig. 1a** shows the results in a cell culture in the exponential growth phase where each cell doubles every 13 hours. Hence, $V(K) \sim 0$ with almost no drift when $N$ is as small as < 50. Genetic drift is expected to increase as $N$ approaches the carrying capacity.

The paradox is exemplified through computer simulations in **Fig. 1b-c**. These panels show the drift pattern in the Haldane model and the standard WF model, respectively, when the populations are growing as the logistic growth model (see Methods). In the WF model, the drift is strong initially but weakened as $N$ increases. The pattern is reversed in the Haldane model. The contrast between the two panels is the paradox.

**Fig. 1 here**

To verify the simulations of **Fig. 1b-c**, we study the growth of yeast populations under laboratory conditions where cells appear to increase following the logistic growth curve. As shown in **Fig. 1d**, the lower portion of the S-curve portrays the exponential growth in a low-density population, while the upper curve indicates a slowdown in growth near the population's carrying capacity. We directly recorded the individual output ($K$) of each yeast cell in the early ($n$ = 25) and late ($n$ = 65) stage of population growth under real-time high-content fluorescence imaging (Methods).

It's evident that $V(K)$ is decoupled from $E(K)$. In the early stage, $E(K)$ is 10.04 with $V(K)$ = 0.60 over five one-hour intervals. In the high-density phase, $E(K)$ decreases to 2.83 but $V(K)$ increases to 1.09 in four intervals (**Fig. 1e-f**). Most interestingly, during the late stage of population growth, $V(\Delta K)$ (adjusted to the four-hour time span for comparison; see Methods) indeed increases as $N$ increases (**Supplementary Table 3**). At the last time point, the population is closest to the carrying capacity and $V(\Delta K)$ experiences a substantial leap from around 1.12 to 1.82. Overall, the variance in progeny number $V(K)$ in a high-density population is always greater that in a low-density population (**Fig. 1g**). Indeed, $V(K)$ may exhibit an increase with the increase in $N$, sometimes outpaces the latter.

Measurements of these two growth phases demonstrate the "paradox of changing $N$" by showing i) there is almost no drift when $N$ is very small; ii) as $N$ increases, $V(K)$, hence, genetic drift would also increase. Furthermore, $V(K)$ may increase sharply when $N$ is close to the carrying capacity. This trend appears to echo the field observations by Coltman, et al. (1999) who show that $V(K)$ increases more than 5-fold when $N$ increases only 2.5-fold in the United Kingdom (UK) population of Soay rams (*Ovis aries*).

*2. The density-dependent Haldane (DDH) model – A first attempt at the Generalized Haldane model*

To explain the paradox of changing $N$, the model has to track and regulate $N$. Ironically, although the branching process can track the changes, the process has been deemed unsuitable for population genetics. This is because $N_t = N_0\, E(K)\text{^}t$ would drive $N$ to approach either zero or infinity as $t$ becomes very large. Indeed, Haldane's original model reset $N'$ to $N$ in each generation. Any generalized Haldane (GH) model must take advantage of this feature by regulating $N$ via $E(K)$, as attempted in this section.

The "paradox of changing $N$" can be explained by $N_e = N/V(K)$ if the numerator and denominator move in the same direction. Most interestingly, $V(K)$ may increase more than $N$ itself. This has been shown when $V(K)$ is close to zero while $N$ is growing exponentially. In addition, when $N$ is very near the carrying capacity, even a small increase in $N$ would intensify the competition and inflate $V(K)$. This is shown in the yeast experiment of **Fig. 1d** and supported by Coltman, et al. (1999) field study of ram populations in UK. Thus, the relationship between $N$ and $V(K)$ may realistically lead to this paradox.

We now extend the Haldane model by incorporating density-dependent regulation of $N$ in order to define the conditions of this paradox. The model developed here is on an ecological, rather than an evolutionary time scale. The simplest model of $N$ regulation is the logistic equation:

$$\frac{dN}{dt} = rN\frac{C_k - N}{C_k} \quad (4)$$

with $C_k$ being the carrying capacity. In the ecological time scale, a changing $N$ would mean that the population is departing from $C_k$ or moving toward it (**Fig. 2a**). (As will be addressed in Discussion, changing $N$ in the WF model is depicted in **Fig. 2b** whereby $N$ is at $C_k$. $C_k$ may evolve too, albeit much more slowly in an evolutionary time scale.) Here, we consider changes in $N$ in the ecological time scale. Examples include the exponential growth when $N \ll C_k$, seasonal fluctuation in $N$ (Frankham 1995; Charlesworth 2009) and competition modeled by the Lotka-Volterra equations (Bomze 1983).

**Fig. 2 here**

Details of the DDH model are given in the Supplementary Information. A synopsis is given here: We consider a non-overlapping haploid population with two neutral alleles. The population size at time $t$ is $N_t$. We assume that expected growth rate $E(K)$ is greater than 1 when $N < C_k$ and less than 1 when $N > C_k$, as defined by Eq. (5) below:

$$E(K_t) = \frac{E(N_{t+1})}{N_t} = \frac{r+1}{1 + r\left(\frac{N_t}{C_k}\right)^z} \quad (5)$$

The slope of $E(K)$ vs. $N$ (i.e., the sensitive of growth rate to changes in population size), as shown in **Fig 3a**, depends on $z$. To determine the variance $V(K)$, we assume that $K$ follows the negative binomial distribution whereby parents would suffer reproduction-arresting injury with a probability of $p_t$ at each birthing (Supplementary Information). Accordingly, $V(K)$ can then be expressed as

$$\frac{V(K_t)}{E(K_t)} = \frac{1-p_t}{p_t} = a(N_t)^b \quad (6)$$

By Eq. (6), the ratio of $V(K)/E(K)$ could be constant, decrease or increase with the increase of population size. With $E(K)$ and $V(K)$ defined, we could obtain the effective population size by substituting Eq. (5) and Eq. (6) into Eq. (3).

$$N_e = \frac{N_t E^2(K_t)}{V(K_t)} = \frac{N_t(r+1)}{a(N_t)^b\left[1 + r\left(\frac{N_t}{C_k}\right)^z\right]} \quad (7)$$

Eq. (7) presents the relationship between effective population size ($N_e$) and the population size ($N$) as shown in **Fig. 3**. The density-dependent $E(K)$ could regulate $N$ with different strength (**Fig. 3a**). The steeper the slope in **Fig. 3a**, the stronger the regulation.

**Fig. 3 here**

The main results of this study are depicted in **Fig. 3b**. First, with no or little regulation of $N$, $N_e$ and $N$ are strongly correlated. The green dashed lines portray the conventional view of decreasing drift with increasing $N$. Second, under a particular strength of $N$ regulation, the red line shows no dependence of $N_e$ on $N$, meaning that genetic drift is independent of $N$. Finally, as $N$ becomes strongly regulated, $N_e$ and $N$ would be negatively correlated as the blue dashed line shows. This trend is the paradox of changing $N$.

In summary, genetic drift effect can indeed decrease, increase or stay nearly constant as the population size increases, depending on the strength of $N$ regulation near $C_k$. We further note that the $V(K)/E(K)$ ratio of Eq. (7) can be independent (**Fig. 3c**), positively dependent (**Fig. 3d**), or negatively dependent (**Fig. 3e**) on $N$. Interestingly, the results of **Fig. 3b** are nearly identical when the $V(K)/E(K)$ ratio changes (**Supplementary Figs 1-3**). These results show that the strength of genetic drift depends on the ecology that governs $E(K)$, $V(K)$ and $N$.

The paradox of changing $N$ may not be an exception but a common rule in natural populations. Note that the WF approximation yielding Eq. (2) assumes nearly constant $N$. The assumption would imply very strong $N$ regulation near $C_k$, which is precisely the condition leading to the paradox of changing $N$ (**Fig. 3**). Coltman et al. (1999)'s observations of the reproduction in rams is such a case.

## II. The paradox of genetic drift in sex chromosomes

Since the relative numbers of Y, X and each autosome (A) in the human population are 1:3:4, the WF model would predict the genetic diversity ($\theta_Y$, $\theta_X$ and $\theta_A$, respectively) to be proportional to 1:3:4. In a survey of human and primate genetic diversity, $\theta_Y$ is almost always

less than expected as has been commonly reported (Hammer, et al. 2001; Wang, et al. 2014; Wilson Sayres, et al. 2014; Makova, et al. 2024). The low $\theta_Y$ value has been used to suggest that human Y chromosomes are under either positive or purifying selection (Wang, et al. 2014; Wilson Sayres, et al. 2014; Makova, et al. 2024). Similarly, the reduced diversity X-linked genes was interpreted as a signature of positive selection (Pan, Liu, et al. 2022).

As pointed out above, under-estimation of genetic drift is a major cause of over-estimation of selective strength. Our goal is hence to see if genetic drift of different strength between sexes can account for the observed genetic diversities. Details will be presented in Wang et al. (2024). Below is a synopsis.

Let $V_m$ and $V_f$ be the $V(K)$ for males and females respectively and $\alpha' = V_m/V_f$ is our focus. (We use $\alpha'$ for the male-to-female ratio in $V(K)$ since $\alpha$ is commonly used for the ratio of mutation rate (Miyata, et al. 1987; Makova and Li 2002).) The three ratios, $\theta_Y/\theta_X$ (denoted as $R_{YX}$), $\theta_Y/\theta_A$ ($R_{YA}$) and $\theta_X/\theta_A$ ($R_{XA}$), could be expressed as the functions of $\alpha'$, which incorporate the relative mutation rates of autosomes, X and Y (Supplementary Information). These relative mutation rates can be obtained by interspecific comparisons as done by Makova and Li (2002).

We note that there are many measures for the within-species genetic diversity, $\theta = 4N_e\mu$. Under strict neutrality and demographic equilibrium, these measures should all converge. In the neutral equilibrium, the infinite site model dictates the frequency spectrum to be $\xi_i = \theta/i$, where $\xi_i$ is the number of sites with the variant occurring $i$ times in $n$ samples. Since every frequency bin is a measure of $\theta$, different measures put different weights on the $i$-th bin (Fay and Wu 2000; Fu 2022). While $\pi$, the mean pairwise differences between sequences, is most commonly used in the literature, we use several statistics to minimize the possible influences of selection and demography (Wang et al. 2024). In this synopsis, we used the Watterson estimator (Watterson 1975) as the measure for $\theta_Y$ , $\theta_X$ and $\theta_A$ by counting the number of segregating sites.

Using any of the three ratios ($\theta_Y/\theta_X$, $\theta_Y/\theta_A$ and $\theta_X/\theta_A$), we can obtain $\alpha'$; for example, $R_{YA} = \theta_Y/\theta_A$ and, hence,

$$R_{YA} = \frac{y}{4}\frac{1+\alpha'}{2\alpha'}$$

where $y$ is the mutation rate of Y-linked sequences relative to autosomal sequences. With rearrangement,

$$\alpha' = \frac{y}{8R_{YA} - y} \qquad (8)$$

Similar formulae of $\alpha'$ can be obtained for $R_{YX}$ and $R_{XA}$ but the accuracy for estimating $\alpha'$ is highest by $R_{YA}$ whereas $R_{XA}$ is the least accurate for this purpose (Supplementary Information).

**Table 2 here**

**Table 2** presents the $\alpha'$ estimates in chimpanzees and bonobos. This is part of a general survey in mammals with a strong emphasis on primates and humans (Wang et al. 2024). It is almost always true that $\alpha' > 5$ in primates. Sources of data used are also given in **Table 2**. As shown for chimpanzees, $\alpha'$ is often far larger than 5, above which the resolution is very low as can be seen in Eq. (8). Note that, when $R_{YA}$ is under-estimated and approaching $y/8$, $\alpha'$ would increase rapidly when the denominator is close to 0. That is why $\alpha'$ often becomes infinity (**Supplementary Fig. 5**). The estimated $\alpha'$ (MSE) in **Table 2** alleviates the problem somewhat by using all three ratios to calculate the mean square error (MSE) (Supplementary Information).

Among primates surveyed, bonobo is the only exception with $\alpha' < 1$. While chimpanzees and bonobos are each's closest relatives, their sexual behaviors are very divergent (de Waal 1995; De Waal and Lanting 2023). With unusually strong matriarchs, bonobo society seems to stand out among primates in sexual dominance. The $\alpha'$ value is important in behavioral studies and, particularly, in primatology and hominoid research. If 0.1% of males sire 100 children and the rest have the same $K$ distribution as females, $V_m/V_f$ would be ~ 10. Such outlier contribution is the equivalent of super-spreaders in viral evolution which may easily be missed in field studies. In this brief exposition, we highlight again that $V(K)$ or, more specifically, $V_m$ and $V_f$ are key to genetic drift. The differences in drift strength among X, Y and autosomes constitute a curious (but not immediately recognizable) paradox for the WF models as will be addressed in Discussion.

## III. The paradox of genetic drift under selection

Genetic drift operates on neutral as well as non-neutral mutations. Let us assume a new mutation, M, with a frequency of $1/N$ is fixed in the population with the probability of $P_f$. Fisher (1930) first suggested that the fixation probability of an advantageous mutation should increase in growing populations, while decrease in shrinking populations. If there is no genetic drift, a beneficial mutation will always be fixed and $P_f = 1$. In the WF model, it is well known that $P_f \sim 2s$ for a new advantageous mutation, with fitness gain of $s$, while population size is large (Haldane 1932; Kimura 1962). This seems paradoxical that the determinant of genetic drift, $1/N_e$, does not influence $P_f$ (Otto and Whitlock 1997; Lanfear, et al. 2014).

**Fig. 4 here**

In the Method section, we show that $P_f$ under the Haldane model is given by

$$P_f = \sum_{t=0}^{\infty} \left[1 - \left(u_M(t)\right)^{M_0}\right] \left[\left(u_W(t)\right)^{W_0} - \left(u_W(t-1)\right)^{W_0}\right] \quad (9)$$

where $M_0$ and $W_0$ are the initial number of mutant allele (M allele) and wildtype allele (W allele), with $N = M_0 + W_0$. We note that $u_M(t)$ and $u_W(t)$, respectively, are the extinction probabilities of M allele and W allele by generation $t$ with the initial number of alleles of 1. While obtaining the direct analytical solution of Eq. (9) may not be feasible, we could obtain its numerical solution due to its convergence as $t$ increases (see Methods). The accuracy of the numerical solution from Eq. (9) is confirmed through simulation (**Supplementary Fig. 4**). Moreover, with the aid of the WF model (see Supplementary Information), we could obtain the approximation of Eq. (9) as follows.

$$P_f = \frac{2s}{V(K)} \quad (10)$$

We verify Eq. (10) by both the numerical solution from Eq. (9) and simulations based on the branching process (**Supplementary Fig. 4**). The fixation probabilities obtained by numerical solution vs. those inferred from Eq. (10) are shown in **Fig. 4**. The salient feature is that the fixation probability of $2s$, as in the classical formula, would be a substantial over-estimate when $V(K)$ is larger than $E(K)$. Eq. (10) is sufficiently accurate as long as $N \geq 50$. When $N$ is as small as 10, the theoretical result is biased. Indeed, at such a low $N$ value, the population is prone to extinction. The DDH model presented above should rectify this deficiency.

The main message of Eq. (10) is that genetic drift under positive selection is influenced by $s$ and $V(K)$ but not by $N$. This independence from $N$ is explicable: when an advantageous mutation increases to a certain level still far below $N$ (depending mainly on $s$), its fixation would be almost certain. This may be a most direct argument against equating genetic drift with $N$, or $N_e$ (which is a function of changing $N$'s).

## Discussion

Genetic drift should broadly include all random forces affecting evolutionary trajectories. Hence, when $N$ is not regulated by the models themselves but supplied externally, some random forces are excluded from the modeling. The DDH model may be a first-generation GH model that incorporates $N$ regulation. We note that the original Haldane model suppresses $N$ fluctuation and is hence a special case of DDH with extremely strong $N$ regulation near $C_k$.

We shall first clarify the first paradox, the "paradox of changing $N$'s". This paradox is in the ecological time scale whereby $N$ is either growing or oscillating around the carrying capacity

(**Fig. 2a**). In this time scale, drift may often (but certainly not always) increase in strength as $N$ increases. While $N$ and $N_e$ are often poorly correlated in WF models(Crow and Kimura 1970; Charlesworth 2009; Lynch, et al. 2016), there is no demonstration how $N$ and $N_e$ can be negatively correlated in the absence of $N$ regulation. Nevertheless, the WF models do work well in the evolutionary time scale (**Fig. 2b**). For example, by the PSMC model (Li and Durbin 2011), both the European and Chinese populations experience a severe bottleneck 10-60 kyr ago. Presumably, various environmental forces may have reduced $C_k$ drastically. Averaged over the long-time span, $N$ should be at or near $C_k$. In short, the N value in the WF models is both N and $C_k$ at the evolutionary time scale. It should also be mentioned that the DDH model is distinct from the concept of "genetic draft" that involves selection and hitchhiking (Gillespie 2000, 2001).

The second paradox of sex-dependent drift is about different $V(K)$'s between sexes (generally $V_m > V_f$) but the same $E(K)$ between them. In the conventional models of sampling, it is not clear what sort of biological sampling scheme could yield $V(K) \neq E(K)$, let alone two separate $V(K)$'s with one single $E(K)$. Mathematically, given separate $K$ distributions for males and females, it is unlikely that $E(K)$ for the whole population could be 1, hence, the population would either explode in size or decline to zero. In short, $N$ regulation has to be built into the genetic drift model as the GH model does to avoid this paradox.

The third paradox of genetic drift is manifested in the fixation probability of an advantageous mutation, $2s/V(K)$. As explained above, the fixation probability is determined by the probability of reaching a low threshold that is independent of $N$ itself. Hence, the key parameter of drift in the WF model, $N$ (or $N_e$), is missing. This paradox supports the assertion that genetic drift is fundamentally about $V(K)$ with $N$ being a scaling factor. Note the absence of genetic drift at all $N$'s when $V(K) = 0$.

The fourth paradox is about the multi-copy gene systems such as viruses and rRNA genes covered in the companion study (Wang, et al. 2024). These systems evolve both within and between hosts. Given the small number of virions transmitted between hosts, drift is strong in both stages as shown by the Haldane model (Ruan, Luo, et al. 2021; Ruan, Wen, et al. 2021; Hou, et al. 2023). Therefore, it does not seem possible to have a single effective population size to account for the genetic drift in two stages with very different biological processes. The inability to deal with multi-copy gene systems may explain the difficulties in accounting for the SARS-CoV-2 evolution (Deng, et al. 2022; Pan, Liu, et al. 2022; Ruan, et al. 2022; Hou, et al. 2023; Ruan, et al. 2023).

As the domain of evolutionary biology expands, many new systems do not have definable populations that fit the criteria of WF populations (such as panmixia in mating or dispersal). Multi-copy gene systems are obvious examples. Others include domestications of animals and plants that are processes of rapid evolution (Diamond 2002; Larson and Fuller 2014; Purugganan 2019; Chen, Yang, et al. 2022; Pan, Zhang, et al. 2022; Wang, et al. 2022). Due to

the very large $V(K)$ in domestication, drift must have played a large role. Somatic cell evolution is another example with "undefinable" genetic drift (Wu, et al. 2016; Chen, et al. 2017; Chen, et al. 2019; Ruan, et al. 2020; Chen, Wu, et al. 2022). The Haldane model is an individual-output model (Chen, et al. 2017) whereby the collection of individuals constitute the "population".

We understand that further modifications of the WF models may account for some or all of these paradoxes. However, such modifications have to be biologically feasible and, if possible, intuitively straightforward. Such possible elaborations of WF models are beyond the scope of this study. We are only suggesting that the Haldane model can be extensively generalized to be an alternative approach to genetic drift. The GH model attempts to integrate population genetics and ecology and, thus, can be applied to genetic systems far more complex than those studied before. The companion study is one such example.

## Methods

**Cell culture and image analysis.** NIH3T3 cells, a fibroblast cell line that was isolated from a mouse NIH/Swiss embryo, were stably transfected with the fluorescent, ubiquitination-based cell cycle indicator (Fucci) (Sakaue-Sawano, et al. 2008) plasmid using Lipofectamine 3000 Transfection Reagent (Invitrogen) following the manufacturer's specified instructions. The Fucci-labeled cells exhibited distinct fluorescent signals indicative of G1, S, and G2/M phases, represented by red, yellow, and green, respectively. NIH3T3-Fucci cell was derived from single-cell colony and cultured in DMEM supplemented with 10% Calf Bovine Serum and penicillin/streptomycin. All cells were maintained at 37□ with 5% $CO_2$. Subsequently, the cells underwent extended time-lapse imaging using high-content fluorescence microscopy (PerkinElmer Operetta CLS) equipped with a 10x objective lens. Images were captured hourly over a 100-hour period, and the analysis was conducted using ImageJ (Fiji) (Schneider, et al. 2012) to count the number of cells (**Supplementary Table 1**).

**Yeast strain construction.** Strains were constructed on the genetic background of *Saccharomyces cerevisiae* strain BY4741. A GFP or BFP fluorescent protein expression cassette, under the control of the TDH3 promoter, was inserted into the pseudogene locus YLL017W. Transformations followed a published protocol (Gietz and Schiestl 2007). Transformants were plated on synthetic complete medium without uracil (SC-Ura), and from these, single colonies were selected. Confirmation of replacements was achieved through PCR, and cassette verification was performed using fluorescence microscopy. Subsequently, the constructed strains were cultivated non-selectively in YPD medium (1% Yeast extract, 2% Peptone, 2% D-glucose) at 30 □ on a rotary shaker.

**Estimation of $V(K)$ and $E(K)$ in yeast cells.** To discern division events, even under high

concentration, we conducted co-cultures of the GFP-yeast and BFP-yeast at ratios of 1:1 and 1:25, with the initial cell concentration of 0.1% and 12.5%, respectively. Then the yeast cells were then continuously imaged under high-content fluorescence microscopy (PerkinElmer Operetta CLS) for 10 hours with 1-hour intervals to observe the individual offspring of GFP-yeast. Yeast cells with a distinct offspring number ($K$) within the initial 5 hours (with the high-density group limited to the first 4 hours) were documented (**Supplementary Table 2**). Subsequently, the mean and variance of $K$ for each cell were calculated across various time intervals as follows.

According to the law of total variance,

$$V(I_t) = E[V(I_t|I_{t-1})] + V[E(I_t|I_{t-1})] = V_{t-1,t}E[I_{t-1}] + E^2{}_{t-1,t}V[I_{t-1}]$$

where $I_t$ is the offspring number at time $t$ for an initial single cell (i.e., the total number of progeny cells for a single cell after $t$ hours), as documented in **Supplementary Table 2**. And $E_{t\text{-}1,\,t}$ represents the average of offspring number after single time interval (1 hour here) for each single cell from time $t$-1 to time $t$, while $V_{t\text{-}1,\,t}$ is the corresponding variance. Utilizing the documented total cell count at time $t$ (denoted by $N_t$), we could calculate the average of offspring number for a single yeast cell in a specific time interval, e.g., $E_{t\text{-}1,\,t} = N_t / N_{t\text{-}1}$ and $E[I_{t\text{-}1}] = N_t / N_0$. With some rearrangement,

$$V_{t-1,t} = \frac{V(I_t) - E^2{}_{t-1,t}V[I_{t-1}]}{E[I_{t-1}]} = \frac{V(I_t) - (N_t/N_{t-1})^2 V[I_{t-1}]}{N_{t-1}/N_0}$$

Applying the aforementioned equations, we observed that both $V_{t\text{-}1,\,t}$ (i.e., $V(K)$ within a one-hour interval) and $V(I_t)$ (i.e., the $V(K)$ from 0 to time $t$) for yeast cells under high density (12.5%) are generally greater than the values under low density (**Supplementary Table 3**). This suggests that the variance of offspring number in high-density conditions could be larger than that in a low-density context. It's noteworthy that the estimated value of $V_{t\text{-}1,\,t}$ could be less than zero (**Supplementary Table 3**), implying a very small variance in the number of offspring. Furthermore, the occurrence of negative values for $V_{t\text{-}1,\,t}$ is more frequent under low density than high density, indicating the higher variance of offspring number in high density as previously suggested.

To show the variance of offspring number overtime, we also calculated the change of offspring number within successive one-hour intervals, denoted as $V(I_t - I_{t\text{-}1})$. This value is intricately linked to the variance of offspring overtime. To facilitate the comparison with the total variance from 0h to 4h, we set $V(\Delta K) = 4V(I_t - I_{t\text{-}1})$ to account for the four-hour time span.

**Simulation of genetic drift in the Haldane model and the Wright-Fisher (WF) model**. In both models, interactions between individuals are implicitly included through the

dependency of the average number of offspring on population size, as defined by Eq. (5). This dependency leads to the logistic population growth, reflecting the density-dependent interactions. The initial population size is set to 4, with initial gene frequency (i.e., the frequency of mutant allele) of 0.5. And the carrying capacity is established at 100, 000, with $r$ = 1 and $z$ = 1. In WF model, the population size at next generation is $N_{t+1} = N_t \times E(K_t)$. Note if the calculated value of $N_{t+1}$ is not an integer, we round it to the nearest whole number. The number of mutant alleles follows a binomial distribution, allowing the simulation of gene frequency overtime in WF model. For Haldane model, the number of offspring number is assumed to follow a beta-binomial distribution. The mean of offspring number $E(K_t)$ is obtained from Eq. (5). The variance $V(K_t)$ is obtained from Eq. (6) with parameters $a$ = 1/300 and $b$ = 1.2. Given the mean and variance of the beta-binomial distribution, we simulated the number of mutant alleles and population size overtime and then traced the gene frequency overtime.

**Fixation probability of a new mutation in Haldane model.** Here we obtain the fixation probability of advantageous mutation in Haldane model governed by a branching process. The Haldane model considers a well-mixed population of $N_t$ haploid parents with only two types of alleles (W for wildtype, M for mutant). There is no migration and new mutations in this model. Each individual is assumed to independently reproduce in discrete and non-overlapping generation, $t$ = 0, 1, 2, …. Thus, the number of offspring per allele (also an individual in this haploid population) at any generation is represented by a set of identically and individually distributed random variables. In particular, the numbers of offspring of M and W alleles (denoted as $K_M$ and $K_W$ respectively) can be represented by following distributions.

$$P(K_M = k) = i_k \text{ for } k = 0, 1, 2, \ldots$$
$$P(K_W = k) = j_k \text{ for } k = 0, 1, 2, \ldots$$

For a particular distribution, we can obtain the average number of offspring as follows.

$$E(K_M) = \sum_{k=0} k \times i_k$$
$$E(K_W) = \sum_{k=0} k \times j_k$$

With the selection coefficient of $s$ (0 for neutral mutation, positive for advantageous mutation, and negative for deleterious mutation) of M allele, average offspring number of M allele and W allele will follow the relationship.

$$E(K_M) = E(K_W)(1 + s)$$

Now, the evolution of the number of M allele (denoted as $M_t$) and the number of W allele

(denoted as $W_t$) as time process is a branching process.

To compare Haldane model with WF model (the offspring number follows Poisson distribution with variance and mean equal to 1), we will set $E(K_W) = 1$, $E(K_M) = 1 + s$. And then let the offspring number follow a more general and realistic distribution (including negative binomial distribution (negBin), beta-binomial distribution (betaBin)), which can let the ratio of variance to mean range from 1 to a large number. (For simplicity, we let $V(K_M)/E(K_M) = V(K_W)/E(K_W)$.) Based on the branching process, we obtained the fixation probability M alleles (see Supplementary Information for more details on the derivation).

$$P_f = \sum_{t=0}^{\infty} \left[1 - \left(u_M(t)\right)^{M_0}\right]\left[\left(u_W(t)\right)^{W_0} - \left(u_W(t-1)\right)^{W_0}\right]$$

where

$$u_M(t) = P(M_t = 0|M_0 = 1) = \sum_{k=0}^{\infty} P(M_t = 0|M_{t=1} = k) = \sum_{k=0}^{\infty} i_k (u_M(t-1))^k$$

$$u_W(t) = P(W_t = 0|W_0 = 1) = \sum_{k=0}^{\infty} P(W_t = 0|W_{t=1} = k) = \sum_{k=0}^{\infty} j_k (u_W(t-1))^k$$

Note $u_M(t = 1) = P(K_M = 0) = i_0$, $u_w(t = 1) = P(K_W = 0) = j_0$. And both $\{u_M(t), t = 0, 1, 2, \ldots\}$ and $\{u_w(t), t = 0, 1, 2, \ldots\}$ are a bounded monotonic sequence. Thus, although we cannot obtain the direct analytical solution for $P_f$, we could easily obtain their numerical solution by iteration to the case when both $u_M(t)$ and $u_w(t)$ converge.

**Acknowledgements**

We thank Xionglei He for helpful comments. The work was supported by the National Natural Science Foundation of China (32150006, 32200493, 32293193, 81972691), the Guangzhou Science and Technology Planning Project (2025A04J3499), the National Key Research and Development Projects of the Ministry of Science and Technology of China (2021YFC2301300, 2021YFC0863400).

**Competing interest statement.** Authors declare that they have no competing interests.

**References**

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14:113-124.

Bomze IM. 1983. Lotka-Volterra equation and replicator dynamics: A two-dimensional classification. Biol. Cybern. 48:201-211.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10:195-205.
Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 355:1563-1572.
Chen B, Shi Z, Chen Q, Shen X, Shibata D, Wen H, Wu CI. 2019. Tumorigenesis as the Paradigm of Quasi-neutral Molecular Evolution. Mol. Biol. Evol. 36:1430-1441.
Chen B, Wu X, Ruan Y, Zhang Y, Cai Q, Zapata L, Wu CI, Lan P, Wen H. 2022. Very large hidden genetic diversity in one single tumor: evidence for tumors-in-tumor. Natl. Sci. Rev. 9:nwac250.
Chen Q, Yang H, Feng X, Chen Q, Shi S, Wu CI, He Z. 2022. Two decades of suspect evidence for adaptive molecular evolution-negative selection confounding positive-selection signals. Natl. Sci. Rev. 9:nwab217.
Chen Y, Tong D, Wu CI. 2017. A New Formulation of Random Genetic Drift and Its Application to the Evolution of Cell Populations. Mol. Biol. Evol. 34:2057-2064.
Coltman DW, Smith JA, Bancroft DR, Pilkington J, MacColl AD, Clutton-Brock TH, Pemberton JM. 1999. Density-dependent variation in lifetime breeding success and natural and sexual selection in Soay rams. Am. Nat. 154:730-746.
Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grutzner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. Nature 508:488-493.
Courret C, Chang CH, Wei KH, Montchamp-Moreau C, Larracuente AM. 2019. Meiotic drive mechanisms: lessons from Drosophila. Proc Biol Sci 286:20191430.
Crow JF, Denniston C. 1988. Inbreeding and Variance Effective Population Numbers. Evolution 42:482-495.
Crow JF, Kimura M. 1970. An Introduction to Population Genetics Theory. New York: Harper & Row.
de Waal FB. 1995. Bonobo sex and society. Sci. Am. 272:82-88.
De Waal FB, Lanting F. 2023. Bonobo: The forgotten ape: Univ of California Press.
Deng S, Xing K, He X. 2022. Mutation signatures inform the natural host of SARS-CoV-2. Natl. Sci. Rev. 9:nwab220.
Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. Nature 418:700-707.
Dubuc C, Ruiz-Lambides A, Widdig A. 2014. Variance in male lifetime reproductive success and estimation of the degree of polygyny in a primate. Behav. Ecol. 25:878-889.
Eldon B, Wakeley J. 2006. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics 172:2621-2633.
Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.
Fisher RA. 1930. The genetical theory of natural selection. Oxford, England: Clarendon Press.
Frankham R. 1995. Effective population size/adult population size ratios in wildlife: a review. Genetics Research 66:95-107.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.
Fu YX. 2022. Variances and covariances of linear summary statistics of segregating sites. Theor. Popul. Biol. 145:95-108.
Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2:31-34.
Gillespie JH. 2000. Genetic drift in an infinite population. The pseudohitchhiking model. Genetics 155:909-919.
Gillespie JH. 2001. Is the population size of a species relevant to its evolution? Evolution 55:2161-2169.
Gillespie JH. 1975. Natural selection for within-generation variance in offspring number II. Discrite haploid models. Genetics 81:403-413.
Hagedoorn AL, Hagedoorn AC. 1921. The relative value of the processes causing evolution. In. The Hague: M. Nijhoff The Hague.
Haldane JBS. 1932. The Causes of Evolution. London: Longmans, Green and Company.
Haldane JBS. 1927. A mathematical theory of natural and artificial selection, part V: selection and mutation: Cambridge University Press.
Hallast P, Maisano Delser P, Batini C, Zadik D, Rocchi M, Schempp W, Tyler-Smith C, Jobling MA. 2016. Great ape Y Chromosome and mitochondrial DNA phylogenies reflect subspecies structure and patterns of mating and dispersal. Genome Res. 26:427-439.
Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. 2001. Hierarchical patterns of global human Y-chromosome diversity. Mol. Biol. Evol. 18:1189-1203.
Hartl DL, Clark AG. 1997. Principles of Population Genetics. Sunderland, Massachusetts: Sinauer Associates.
Hasselquist D. 1995. Demography and Lifetime Reproductive Success in the Polygynous Great Reed Warbler. Japanese Journal of Ornithology 44:181-194.
Herman DM, Colwell MA. 2015. Lifetime reproductive success of Snowy Plovers in coastal northern California. The Condor 117:473-481.
Hou M, Shi J, Gong Z, Wen H, Lan Y, Deng X, Fan Q, Li J, Jiang M, Tang X, et al. 2023. Intra- vs. Interhost Evolution of SARS-CoV-2 Driven by Uncorrelated Selection-The Evolution Thwarted. Mol. Biol. Evol. 40.
Kendall MG, Stuart A, Ord JK, Arnold SF. 2006. Kendall's advanced theory of statistics. London: Edward Arnold London.
Kimura M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47:713-719.
Kimura M, Crow JF. 1963. The measurement of effective population number. Evolution:279-288.
Krüger O, Lindström J. 2001. Lifetime Reproductive Success in Common Buzzard, Buteo buteo: From Individual Variation to Population Demography. Oikos 93:260-273.
Kuester J, Paul A, Arnemann J. 1995. Age-related and individual differences of reproductive

success in male and female barbary macaques (Macaca sylvanus). Primates 36:461-476.
Lanfear R, Kokko H, Eyre-Walker A. 2014. Population size and the rate of evolution. Trends Ecol. Evol. 29:33-41.
Larson G, Fuller DQ. 2014. The Evolution of Animal Domestication. Annual Review of Ecology, Evolution, and Systematics 45:115-136.
Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature 475:493-496.
Li W-H. 1997. Molecular evolution. Sunderland: Sinauer Associates Incorporated.
Lindholm AK, Dyer KA, Firman RC, Fishman L, Forstmeier W, Holman L, Johannesson H, Knief U, Kokko H, Larracuente AM, et al. 2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. Trends Ecol. Evol. 31:315-326.
Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. Nat. Rev. Genet. 17:704-714.
Makova KD, Li WH. 2002. Strong male-driven evolution of DNA sequences in humans and apes. Nature 416:624-626.
Makova KD, Pickett BD, Harris RS, Hartley GA, Cechova M, Pal K, Nurk S, Yoo D, Li Q, Hebbar P, et al. 2024. The complete sequence and comparative analysis of ape sex chromosomes. Nature 630:401-411.
Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T. 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harb Symp Quant Biol 52:863-867.
Nam K, Munch K, Hobolth A, Dutheil JY, Veeramah KR, Woerner AE, Hammer MF, Great Ape Genome Diversity P, Mailund T, Schierup MH. 2015. Extreme selective sweeps independently targeted the X chromosomes of the great apes. Proc Natl Acad Sci U S A 112:6413-6418.
Oring LW, Colwell MA, Reed JM. 1991. Lifetime reproductive success in the spotted sandpiper (Actitis macularia) : sex differences and variance components. Behav. Ecol. Sociobiol. 28:425-432.
Otto SP, Whitlock MC. 1997. The probability of fixation in populations of changing size. Genetics 146:723-733.
Pan Y, Liu P, Wang F, Wu P, Cheng F, Jin X, Xu S. 2022. Lineage-specific positive selection on ACE2 contributes to the genetic susceptibility of COVID-19. Natl. Sci. Rev. 9:nwac118.
Pan Y, Zhang C, Lu Y, Ning Z, Lu D, Gao Y, Zhao X, Yang Y, Guan Y, Mamatyusupu D, et al. 2022. Genomic diversity and post-admixture adaptation in the Uyghurs. Natl. Sci. Rev. 9:nwab124.
Purugganan MD. 2019. Evolutionary Insights into the Nature of Plant Domestication. Curr. Biol. 29:R705-R714.
Ribble DO. 1992. Lifetime Reproductive Success and its Correlates in the Monogamous Rodent, Peromyscus californicus. The Journal of Animal Ecology 61:457-468.
Ruan Y, Luo Z, Tang X, Li G, Wen H, He X, Lu X, Lu J, Wu CI. 2021. On the founder effect

in COVID-19 outbreaks: how many infected travelers may have started them all? Natl. Sci. Rev. 8:nwaa246.
Ruan Y, Wang H, Chen B, Wen H, Wu CI. 2020. Mutations Beget More Mutations-Rapid Evolution of Mutation Rate in Response to the Risk of Runaway Accumulation. Mol. Biol. Evol. 37:1007-1019.
Ruan Y, Wen H, He X, Wu CI. 2021. A theoretical exploration of the origin and early evolution of a pandemic. Sci Bull (Beijing) 66:1022-1029.
Ruan Y, Wen H, Hou M, He Z, Lu X, Xue Y, He X, Zhang YP, Wu CI. 2022. The twin-beginnings of COVID-19 in Asia and Europe-one prevails quickly. Natl. Sci. Rev. 9:nwab223.
Ruan Y, Wen H, Hou M, Zhai W, Xu S, Lu X. 2023. On the epicenter of COVID-19 and the origin of the pandemic strain. Natl. Sci. Rev. 10:nwac286.
Sackman AM, Harris RB, Jensen JD. 2019. Inferring Demography and Selection in Organisms Characterized by Skewed Offspring Distributions. Genetics 211:1019-1028.
Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, Osawa H, Kashiwagi S, Fukami K, Miyata T, Miyoshi H, et al. 2008. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell 132:487-498.
Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671-675.
Setchell JM, Charpentier M, Wickings EJ. 2005. Sexual selection and reproductive careers in mandrills (Mandrillus sphinx). Behav. Ecol. Sociobiol. 58:474-485.
Sibly RM, Hone J. 2002. Population growth rate and its determinants: an overview. Philos Trans R Soc Lond B Biol Sci 357:1153-1170.
Silver LM. 1985. Mouse t haplotypes. Annu. Rev. Genet. 19:179-208.
Smith JM, Slatkin M. 1973. The Stability of Predator□Prey Systems. Ecology 54:384-391.
Wang CC, Jin L, Li H. 2014. Natural selection on human Y chromosomes. J Genet Genomics 41:47-52.
Wang X, He Z, Guo Z, Yang M, Xu S, Chen Q, Shao S, Li S, Zhong C, Duke NC, et al. 2022. Extensive gene flow in secondary sympatry after allopatric speciation. Natl. Sci. Rev. 9:nwac280.
Wang X, Ruan Y, Zhang L, Chen X, Shi Z, Wang H, Chen B, Tracy M, Wen H, Wu C-I. 2024. The paradox of extremely fast evolution driven in multi-copy gene systems - A resolution. In: eLife Sciences Publications, Ltd.
Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7:256-276.
Wilson Sayres MA, Lohmueller KE, Nielsen R. 2014. Natural selection reduced diversity on human y chromosomes. PLoS Genet. 10:e1004064.
Wright S. 1969. Evolution and the Genetics of Populations, Volume 2: Theory of gene frequencies: University of Chicago press.
Wright S. 1931. Evolution in Mendelian Populations. Genetics 16:97-159.
Wu CI, Lyttle TW, Wu ML, Lin GF. 1988. Association between a satellite DNA sequence and

the Responder of Segregation Distorter in D. melanogaster. Cell 54:179-189.
Wu CI, True JR, Johnson N. 1989. Fitness reduction associated with the deletion of a satellite DNA array. Nature 341:248-251.
Wu CI, Wang HY, Ling S, Lu X. 2016. The Ecology and Evolution of Cancer: The Ultra-Microevolutionary Process. Annu. Rev. Genet. 50:347-369.
Wysocki D, Jankowiak Ł, Cholewa M, Zyskowski D, Griffin A. 2019. Natal conditions, lifespan and lifetime reproductive success of European blackbirds. Behav. Ecol. 30:1707-1714.

## Figures and Tables

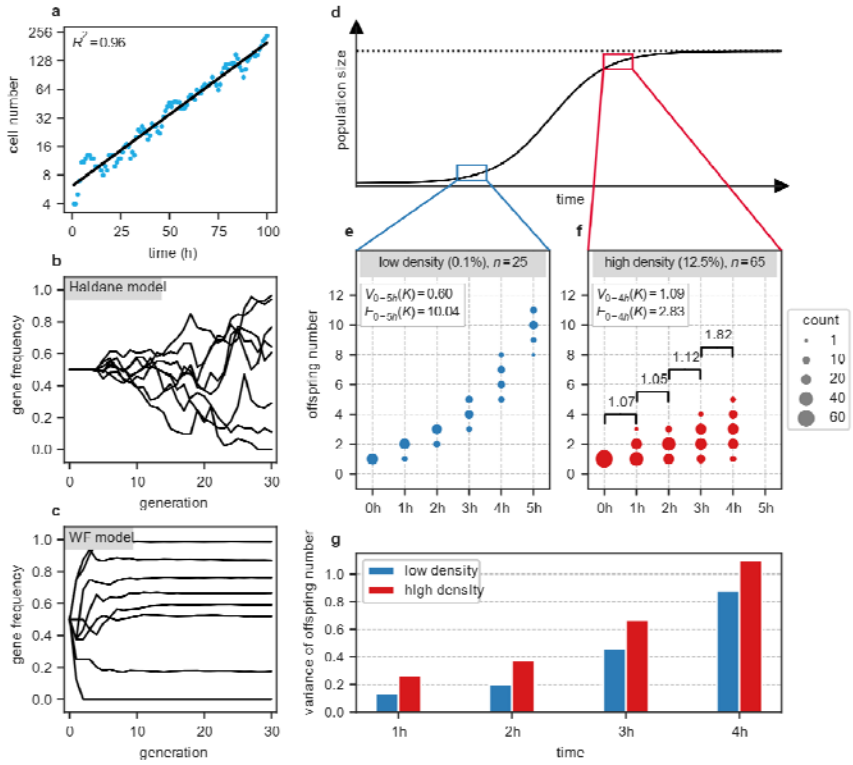


**Fig. 1. The paradox of genetic drift when the population size (*N*) changes**. (**a**) In the laboratory cell culture, nearly all cells proliferate when *N* is very small (interpreted in the next panel). (**b**) The simulation of Haldane model shows little genetic drift in the exponential phase. As in (a), drift may increase due to the heightened competition as the population grows. (**c**) Simulation by the WF model shows a pattern of drift opposite of the Haldane model. (**d**) The patterns of drift at the low and high *N* are analyzed in the framework of the logistic growth model. (**e-f**) Measurements of genetic drift in laboratory yeast populations at low and high density as defined in (d). The progeny number of each cell, *K*, is counted over 4 or 5 intervals as shown by the dots, the sizes of which reflect the cell number. $E(K)$ and $V(K)$ are presented as well. In panel (**f**), the change of offspring number overtime, denoted as $V(\Delta K)$, is shown above the braces. (**g**) The variance of offspring number $V(K)$ increases, observed in (e) and (f), as population size increases.

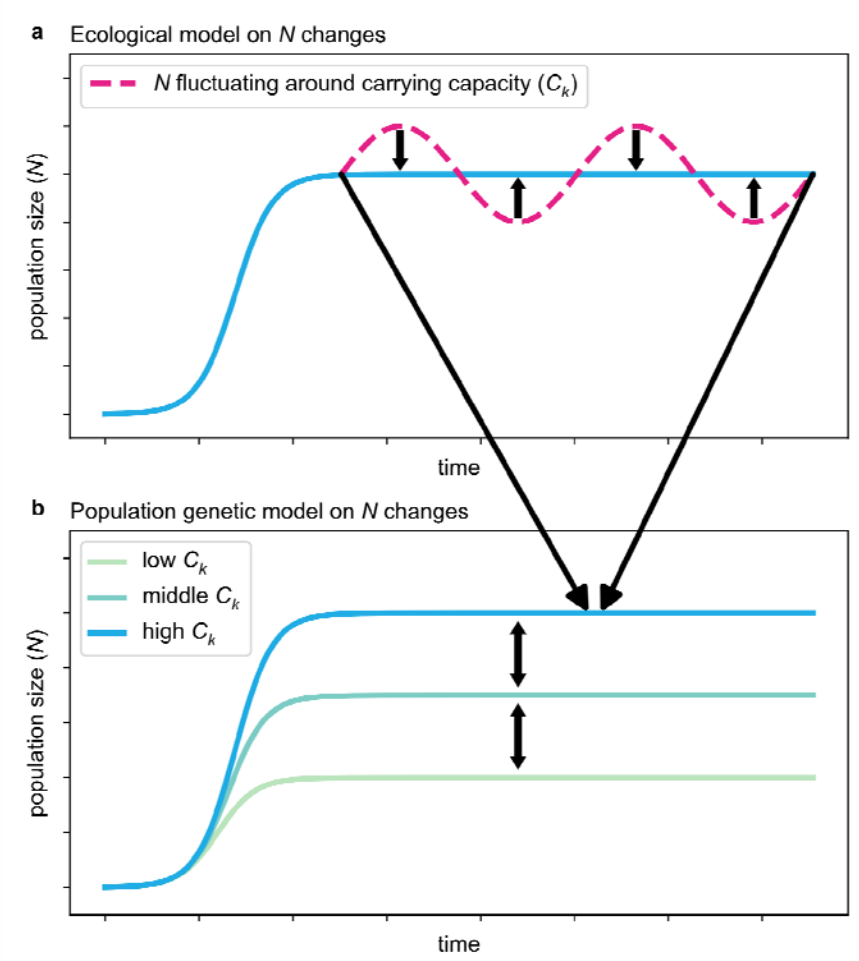


**Fig. 2. The meaning of population size (*N*) changes in ecology vs. in population genetics.** (**a**) In ecology, changing *N* would generally mean a population approaching or departing the carrying capacity. (**b**) In population genetics, a population of size *N* is assumed to be at the carrying capacity, $C_k$. Thus, changes in *N* would mean an evolving $C_k$, likely the consequence of environmental changes. The arrows indicate the disparity in time scale between the two scenarios.

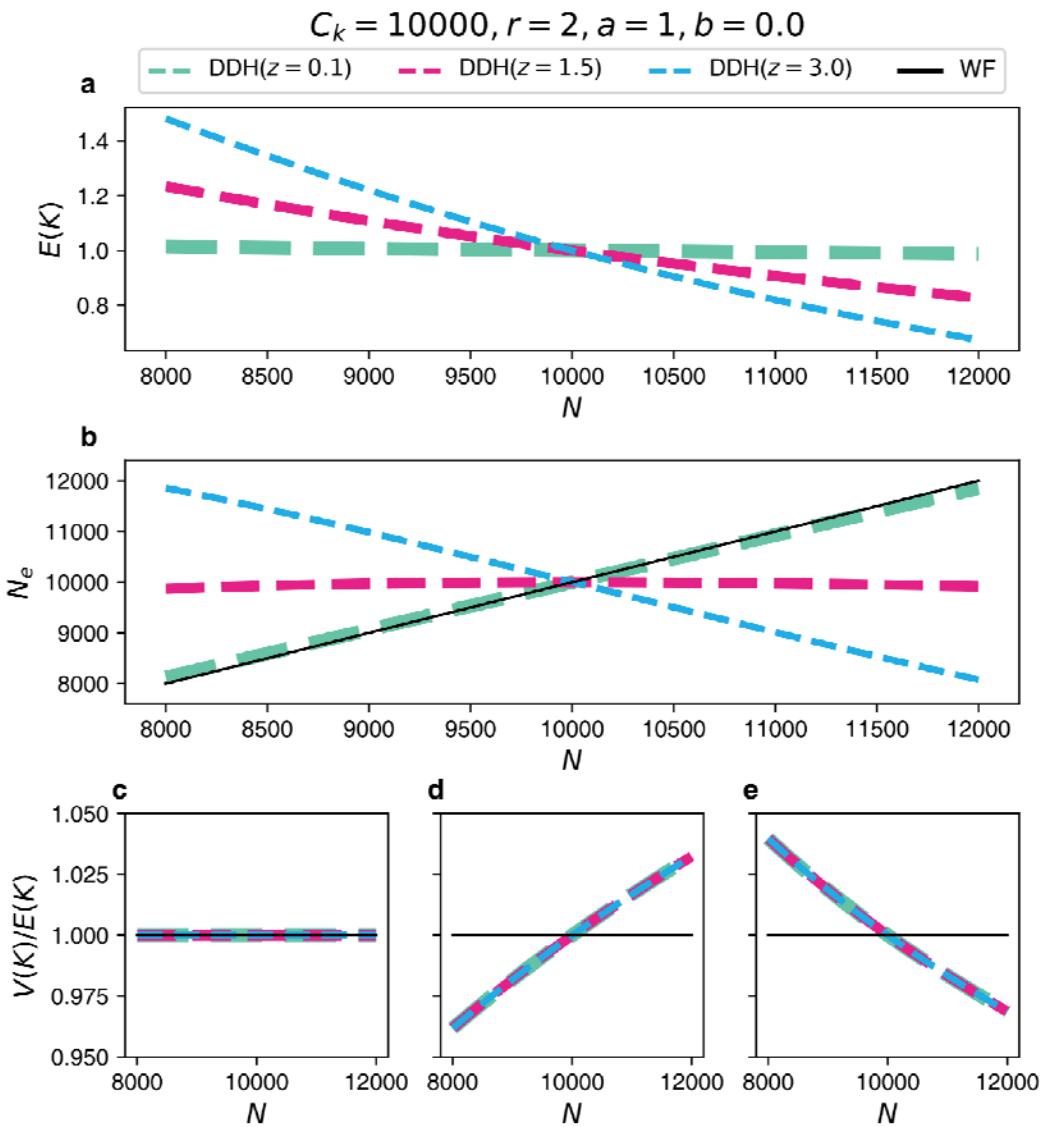


**Fig. 3. Genetic drift as a function of population size in the DDH model.** For all panels, the carrying capacity is $C_k$ = 10,000 and the intrinsic growth rate is $r$ = 2. (**a**) When $N$ increases, $E(K)$ decreases as modeled in Eq. (5). The $z$ value of Eq. (5) (0.1, 1.5 and 3) determines the strength of $N$ regulation, indicated by the slope of $E(K)$ near $C_k$ = 10,000. (**b**) Depending on the strength of $N$ regulation near $C_k$, genetic drift can indeed decrease, increase or stay nearly constant as the population size increases. Thus, the conventional view of $N_e$ being positively dependent on $N$ is true only when the regulation of $N$ is weak (the green line). At an intermediate strength (the red line), $N_e$ is nearly independent of $N$. When the regulation becomes even stronger at z = 3, $N_e$ becomes negatively dependent on $N$. (**c-e**) $V(K)/E(K)$ of Eq. (6)) is shown as a function of $N$. The results of panel (b) are based on a constant $V(K)/E(K)$ shown in panel (c). Interestingly, the results of panel (b) would not be perceptibly changed when $V(K)/E(K)$ varies, as shown in panel (d) and (e).

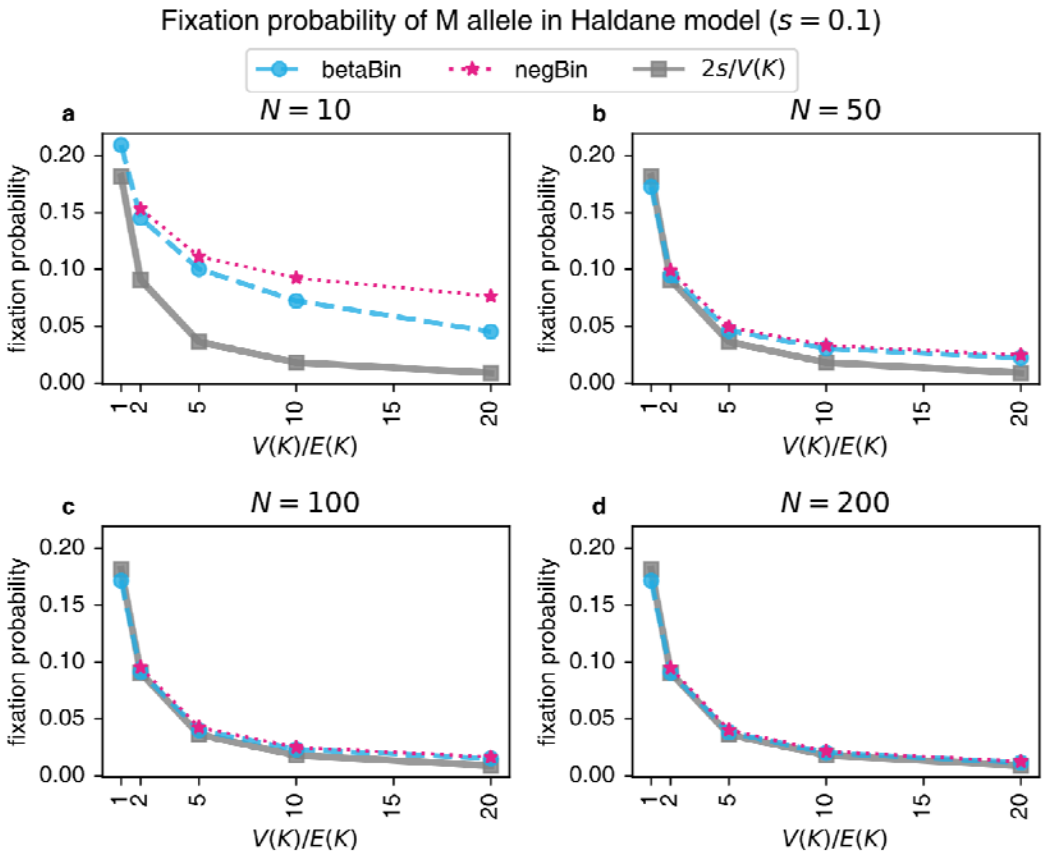


**Fig. 4. Fixation probability of a new advantageous mutation in the Haldane model.** The fixation probabilities of a new advantageous mutation with the selective advantage of $s$ = 0.1 are calculated based on approximate solution from Eq. (9) (i.e., $2s/V(K)$) as well as numerical solution from Eq. (10). The numerical solution from Eq. (10) has been confirmed accurate by simulations (Supplementary Fig. 4). **(a-b)** When $N < 50$, the approximate fixation probability (the gray line) is lower than the simulated values (the color lines) due to population extinction. **(c-d)** By the Haldane model, the expected fixation probability of Eq. (9) is accurate when $N$ reaches 100, as in most natural populations.

**Table 1. Field record of $E(K)$ and $V(K)$ across diverse taxa.**

| Species | Type | Order | Gender | $E(K)$ | $V(K)$ | $V(K)/E(K)$ | Sample Size | Literature |
|---|---|---|---|---|---|---|---|---|
| *Macaca mulatta* | Mammal | Primates | Female | 7.7 | 15.4 | 2.0 | 132 | Dubuc, et al. (2014) |
| | | | Male | 8.7 | 98.7 | 11.3 | 86 | |
| *Mandrillus sphinx* | Mammal | Primates | Female | 4.7 | 19.8 | 4.2 | 51 | Setchell, et al. (2005) |
| | | | Male | 3.5 | 68.2 | 19.5 | 53 | |
| *Macaca sylvanus* | Mammal | Primates | Female | 1.2 | 2.9 | 2.4 | 98 | Kuester, et al. (1995) |
| | | | Male | 1.5 | 3.9 | 2.7 | 81 | |
| *Peromyscus californicus* | Mammal | Rodentia | Female | 4.7 | 7.7 | 1.6 | 16 | Ribble (1992) |
| | | | Male | 4.4 | 13.7 | 3.1 | 20 | |
| *Turdus merula* | Bird | Passeriformes | Female | 4.9 | 23.4 | 4.8 | 80 | Wysocki, et al. (2019) |
| | | | Male | 4.0 | 27.0 | 6.8 | 72 | |
| *Acrocephalus arundinaceus* | Bird | Passeriformes | Female | 7.7 | 53.3 | 6.9 | 82 | Hasselquist (1995) |
| | | | Male | 10.5 | 132.3 | 12.6 | 64 | |
| *Charadrius nivosus* | Bird | Charadriiformes | Female | 1.7 | 5.3 | 3.1 | 105 | Herman and Colwell (2015) |
| | | | Male | 2.2 | 9.0 | 4.1 | 90 | |
| *Actitis macularia* | Bird | Charadriiformes | Female | 5.2 | 26.9 | 5.1 | 66 | Oring, et al. (1991) |
| | | | Male | 3.3 | 17.2 | 5.2 | 103 | |
| *Buteo buteo* | Bird | Accipitriformes | Female | 3.5 | 13.8 | 4.0 | 106 | Krüger and Lindström (2001) |
| | | | Male | 2.7 | 9.3 | 3.4 | 132 | |

**Table 2. Estimation of $V_m/V_f$ in chimpanzee and bonobo.**

| | $\theta_Y$ ($\times 10^{-3}$) | $\theta_X$ ($\times 10^{-3}$) | $\theta_A$ ($\times 10^{-3}$) | $\alpha'$ (MSE) |
|---|---|---|---|---|
| **Chimpanzee** | | | | |
| Western | 0.034 | 0.533 | 1.623 | 4.98 |
| Eastern | 0.144 | 1.044 | 1.446 | 25.90 |
| Nigeria-Cameroon | 0.213 | 0.818 | 0.703 | Inf [a] |
| **Bonobo** | 0.492 | 0.446 | 0.867 | 0.57 |

Data for counting the segregating sites are from Hallast, et al. (2016) and Nam, et al. (2015).

$\theta_Y$, $\theta_X$, $\theta_A$: genetic diversity for Y chromosome, X chromosome, and Autosomes (A), measured by Watterson estimator (Watterson 1975).

$\alpha'$ (MSE): optimal estimate for $\alpha'$ based on the minimum mean square error (MSE).

[a] Inf (infinity) designates cases when the estimated $\alpha'$ based on MSE is greater than 1000 (i.e., $V_m \gg V_f$).