Fixation when *N* and *s* Vary: Classic Approaches Give Elegant New Results

L. M. Wahl[1]
Department of Applied Mathematics, University of Western Ontario, London, ON, Canada N6A 5B7

In this commentary, GENETICS' Associate Editor Lindi Wahl examines R. A. Fisher's 1922 paper "On the dominance ratio" in light of two modern responses to the question of allele fixation probabilities. The two articles that Wahl discusses are published in this month's GENETICS.

IN 1922, an article entitled "On the dominance ratio" appeared in the *Proceedings of the Royal Society of Edinburgh*. The author was R. A. Fisher, a 29-year-old statistician studying crop variation at the Rothamsted Experimental Station in Harpenden, England. The article was a follow-up to Fisher's 1918 article on the statistical effects of Mendelian inheritance, but it is in the 1922 article that Fisher reveals most clearly his uncanny prescience of the questions and techniques that would dominate theoretical population genetics over the next century (Fisher 1918, 1922).

In section 2 of that article, Fisher considers the survival of rare "mutant genes" and introduces what we now call a branching process to address this question. The succinctness with which Fisher explains his approach was typical of the author and does not seem overly concise to today's reader. Fisher's brevity, however, is astounding when we consider that his 1922 exposition was the *first* application of branching processes to any field of science after Galton and Watson's development of these techniques to explain the extinction of English surnames (Watson and Galton 1874). [The underlying mathematics was later rediscovered for application to nuclear chain reactions; this rediscovery was necessary, according to Fisher, because physicists considered him "an ignorant country bumpkin" (Gale 1990, p. 114, citing personal communication with Fisher).]

doi: 10.1534/genetics.111.131748
[1]Address for correspondence: Department of Applied Mathematics, Middlesex 255, University of Western Ontario, London, ON, Canada N6A 5B7. E-mail: lwahl@uwo.ca

Similarly, in section 3 of the same article (Fisher 1922), a half-page derivation is provided for a heat equation, that is, a constant-coefficient diffusion equation, describing the time evolution of the gene frequency distribution. This was the first application of a diffusion process to population genetics, and Fisher, known for his distrust of methods borrowed from other disciplines, derives the approach from first principles without appealing to parallel work in stochastic processes (see Feller 1951).

In the decades of work that followed Fisher's 1922 article, branching processes and diffusion approximations became the two classic approaches for estimating fixation probabilities, that is, the probability that a segregating allele is ultimately carried by all individuals in a population. Most notably, eccentric English biologist J. B. S. Haldane, presumably during breaks from his studies of human physiology through dangerous self-experimentation (Crow 1992), used branching processes to derive the well-known probability that a rare, slightly beneficial allele will fix (Haldane 1927). Later, Motoo Kimura applied the diffusion approach to derive a more general expression for the fixation probability (Kimura 1955). In the second appendix of his well-known chapter (Feller 1951), Feller provides a formal "passage" between the two approaches by recasting the generating function $f(x)$ as a characteristic function, $f(e^{iz})$.

How have both of these "rival" approaches remained current as the field has advanced over decades? Diffusion approximations are far more powerful: a diffusion equation can predict the dynamic frequencies, and eventual fixation, of beneficial, neutral, or deleterious alleles, starting from any initial frequency in the population. Yet, the assumptions underlying diffusion approximations are more difficult to understand, and the technique nearly always requires a numerical approach (number-crunching by computer) to give predictive results. In contrast, branching processes are limited in their applicability: they can be used only to explore situations in which the mutation of interest is beneficial and the mutant allele is initially rare. In these limited

circumstances, however, a branching process is easily formulated and elegant, and can provide compact approximations such as Haldane's famous 2*s* result (Haldane 1927). A branching process often has the advantage of being the *simplest possible model*—no more complex than necessary for the question of interest—but not easily generalizable to other questions.

It would be difficult to imagine a better illustration of the strength, and relevance, of these two approaches than two articles published in this month's issue (Uecker and Hermisson 2011; Waxman 2011). Both articles address the same question about fixation probabilities, that is: How is fixation affected if both the strength of selection, s, and the population size, N, vary in time? As described in more detail by the authors themselves, these two concerns each have a rich history in the literature. Anecdotally, to quote Feller on the assumption of constant N, "essential features of the whole mathematical theory depend on this assumption. Dropping it will lead us to an entirely new theoretical model" (Feller 1951, p. 228). Likewise, Fisher himself overturned the assumption of constant selective pressures in an experimental study of Oxfordshire moths, concluding that "natural populations . . . are affected by selective action varying from time to time in direction and intensity" (Fisher and Ford 1947, p. 171).

The two articles published in this issue address simultaneous changes in selective pressure and population size, and their impact on fixation. Both contributions allow for arbitrary changes in population size; Waxman treats an arbitrary effective population size, $N_e(t)$, while Uecker and Hermisson allow for time-varying birth and death rates, $b(t)$ and $d(t)$. Similarly, the authors of both articles allow the strength of selection to follow an arbitrary time course $s(t)$. [This is in contrast with the well-studied scenario in which s is allowed to fluctuate, but is sampled from an underlying distribution that does not vary in time (Kimura 1962).] To predict fixation probabilities, Waxman takes a diffusion approach, while Uecker and Hermisson chiefly use branching processes. The strengths and limitations of the articles are a textbook lesson in applying the two approaches: Waxman has no restrictions on the sign or size of s, nor on the initial frequency of the allele of interest; Uecker and Hermisson treat only beneficial mutations that are initially rare.

Waxman's approach is not entirely unrestricted, however. An extremely elegant expression for the fixation probability (equation 4 in Waxman 2011) is provided, assuming that the product $S = s(t)N_e(t)$ may vary arbitrarily up to some time T, but thereafter remains constant. The expression for the fixation probability ultimately depends on the entire probability distribution for the allele frequency at this time T, which must be obtained through simulation or through the numerical solution of a backward diffusion equation. Thus, another way to interpret this contribution is that it provides a simple closed-form expression for the fixation probability, given the probability distribution of the allele frequency at an arbitrary time, and constant S thereafter.

In contrast, Uecker and Hermisson have a more restrictive focus, but are able to derive both fixation probabilities and passage times, that is, expected times required for the allele to reach a given frequency. The expression for the fixation probability provided by these authors depends on the entire history ($t = 0$ to ∞) of both $N_e(t)$ and $s(t)$ (equation 16b in Uecker and Hermisson 2011). There is no assumption that N_e or s reach asymptotically constant values, but integrals involving these functions must be obtained, such that, for most cases of interest, numerical techniques will ultimately be required as well. In the supporting information for their article, Uecker and Hermisson also apply the classic diffusion approximation to the case of varying s and N, demonstrating that a Kolmogorov backward equation holds (with time-varying coefficients), and using a clever Ansatz to allow for an approximate solution when the fixation probability is small.

In the decades since Fisher's early articles were published, interest in the fixation process has rarely waned, in part because fixation underpins larger issues such as molecular evolution, standing genetic variation, and adaptation. How do the contributions that follow shed light on these inquiries? To break new ground with arbitrary $s(t)$ and $N(t)$, the authors use models that are, in other respects, highly simplified. For example, Uecker and Hermisson use a haploid model, while Waxman models diploid semidominance, that is, heterozygote and homozygote advantage of $1 + s$ and $1 + 2s$, respectively. Thus neither model can treat arbitrary dominance effects, such as overdominance, which complicate the fixation process, nor can the models address more complex scenarios that maintain standing variation, such as frequency-dependent selection. Nonetheless, the results of these researchers will immediately allow for better estimates of adaptation rates and lay the groundwork for more complex models in the future.

Apart from the textbook clarity in their application of classic theoretical approaches, the articles highlighted here are noteworthy for the sheer simplicity of the question that they address. Although branching processes and diffusion approximations are not standard training for many, the research question here is elegant, relevant, and could be explained by any geneticist: How likely is a new mutation to spread through the population, if the population size and the advantage of the mutant are changing in time? One of the attractions of theoretical population genetics is that simple, tangible questions such as this are still being asked. I encourage you to have a look at the following articles and appreciate for yourself the lucid answers provided by Waxman and by Uecker and Hermisson.

Literature Cited

Crow, J. F., 1992 Centennial: J. B. S. Haldane, 1892–1964. Genetics. 130: 1–6.

Feller, W., 1951 Diffusion processes in genetics, pp. 227–246 in *Proceedings of the Second Berkeley Symposium on Mathematical*

Statistics and Probability, edited by J. Neyman. University of California Press, Berkeley.
Fisher, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52: 399–433.
Fisher, R. A., 1922 On the dominance ratio. Proc. R. Soc. Edinb. 42: 321–341.
Fisher, R. A., and E. B. Ford, 1947 The spread of a gene in natural conditions in a population of the moth *Panaxia dominula* L. Heredity. 1: 143–173.
Gale, J. S., 1990 *Theoretical Population Genetics*. Unwin Hyman, London.
Haldane, J. B. S., 1927 A mathematical theory of natural and artificial selection. V. Selection and mutation. Proc. Camb. Philos. Soc. 23: 838–844.
Kimura, M., 1955 Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harb. Symp. Quant. Biol. 20: 33–53.
Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Genetics 47: 713–719.
Uecker, H. and J. Hermisson, 2011 On the fixation process of a beneficial mutation in a variable environment. Genetics 188: 915–930.
Watson, H. W., and F. Galton, 1874 On the probability of the extinction of families. J. Anthropol. Inst. G B Irel. 4: 138–144.
Waxman, D., 2011 A unified treatment of the probability of fixation when population size and the strength of selection change over time. Genetics 188: 907–913.