

Opinion

Mayr, mathematics and the study of evolution

James F Crow

Address: Genetics Laboratory, University of Wisconsin, 425 G Henry Mall, Madison, WI 53706-1574, USA. Email: jfcrow@wisc.edu

Published: 23 February 2009

Journal of Biology 2009, **8**:13 (doi:10.1186/jbiol117)

The electronic version of this article is the complete one and can be found online at <http://jbiol.com/content/8/2/13>

© 2009 BioMed Central Ltd

Abstract

In 1959 Ernst Mayr challenged the relevance of mathematical models to evolutionary studies and was answered by JBS Haldane in a witty and convincing essay. Fifty years on, I conclude that the importance of mathematics has in fact increased and will continue to do so.

In 1959 Ernst Mayr (Figure 1) flung down the gauntlet [1] at the feet of the three great population geneticists RA Fisher, Sewall Wright and JBS Haldane (Figure 2): "But what, precisely," he said, "has been the contribution of this mathematical school to the evolutionary theory, if I may be permitted to ask such a provocative question?" His skepticism arose in part from the fact that the mathematical theory at the time had little to say about speciation, Mayr's major interest. But his criticism was more broadly addressed to the utility of the entire approach. A particular focus was the simplification that he called "beanbag genetics", in which "Evolutionary genetics was essentially presented as an input or output of genes, as the adding of certain beans to a beanbag and the withdrawing of others." [1].

Mayr was, however, criticizing textbook simplifications, rather than the actual work of the three pioneers. Far from treating gene frequency changes as analogous to the consequence of beans jostling at random in a bag, both Fisher and Wright considered gene interactions in detail. Fisher (Figure 2a) showed that, despite interactions between genes, natural selection acts on the additive component of the genetic variance. It is as if nature were familiar with least squares. The beanbag criticism was particularly inappropriate for Wright (Figure 2b), who specifically devised his 'shifting balance' theory as a way for a population to go

from one harmonious gene combination (Mayr would say "integrated genotype") to another when intermediates were disadvantageous.

Who was to answer Mayr's criticism? Fisher was already dead, and in any case preferred attack to defense, and Wright was too gentle - though admittedly not always when Mayr was involved: returning from Italy where he had received the prestigious Balzan Prize in 1984, Wright told me that the value of the prize was considerably diminished when he discovered that Mayr had won it the year before. In the event, however, it was Haldane (Figure 2c) who took up the challenge. And he did it with flair and gusto. The result was "A defense of beanbag genetics" [2]. This was Haldane at his best - witty, spirited, informed, interesting and convincing.

But the larger question remains: what indeed has been the contribution of mathematical theory to evolution? Mathematics is not central to evolution in the way it has been in theoretical physics. Solid advances have been made without using mathematics, much being due to Mayr himself [3]. And these continue. Yet, I shall argue that mathematical ideas have made important, and often essential, contributions, and still do. Many concepts that are now established were arrived at mathematically, although their origins have since been forgotten.

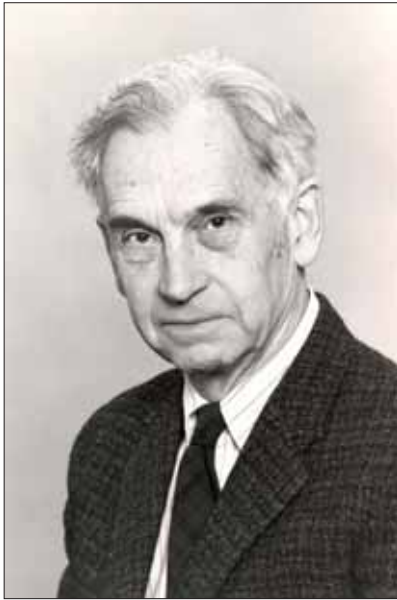


Figure 1
Ernst Mayr (1904-2005). Photograph reproduced with permission from the Archives of the Ernst Mayr Library of the Museum of Comparative Zoology, Harvard University.

For example, the idea that polymorphisms become stabilized in populations because heterozygotes are at an advantage is now found in elementary textbooks, but Fisher was the first to formulate it. Loss of heterozygosity with inbreeding is also textbook knowledge, but it was not clear until Wright developed the theory and invented a simple algorithm for quantifying it. Similarly, the idea that the impact of mutation on the population depends on the mutation rate rather than the magnitude of the mutant effect is now taken for granted, but that was not known until Haldane showed it mathematically. One final example is the inheritance of the ABO blood groups, which was in doubt from the time of their discovery at the turn of the twentieth century until Bernstein's mathematical population analysis in 1924 [4]. All of these applications used only elementary methods, and they must have been known to Mayr. Often, concepts that were developed mathematically were later explained in intuitive, non-mathematical ways. HJ Muller was particularly inventive in finding such explanations. But the mathematical derivation usually came first. It's a lot easier to find an intuitive explanation when you already know the answer.

Ironically, Mayr himself unwittingly provided an especially compelling argument for mathematical analysis. His theory of "genetic revolutions" assumed that from a well integrated population, genetic drift in a small founder offshoot will

sometimes produce a population with a new set of genotypes integrated in a new way. Intuitively, a small founder population seemed a particularly unlikely place to find a new favorable gene combination, and this was indeed shown to be the case in a very detailed mathematical analysis by Barton and Charlesworth [5]. If Mayr had had more respect for mathematical population genetics, he never would have made what most theorists regard as the mistake of proposing that small founder populations are a likely source of major evolutionary changes by genetic drift.

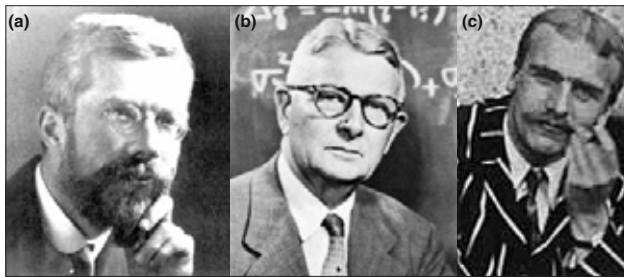
Recent mathematical work has gone well beyond that of the three pioneers. Partly this is due to skilled mathematicians entering the field and bringing new techniques with them; especially noteworthy are stochastic processes. Second, and perhaps more important, is the extensive use of computers. Often you can use a computer to get by without deep mathematical knowledge. An additional influence is the explosive growth of molecular data, which lend themselves to mathematical treatment. In the first half of the twentieth century, population genetics and evolution had a beautiful theory, but there were very limited opportunities to apply it. Now the situation is reversed. Molecular data accumulate too fast to be assimilated.

What are some of the newer developments in evolution that are owed to mathematical theory? Here are a few.

Neutral theory, molecular clocks and selective sweep

One striking result in the post-Mayr period was Motoo Kimura's neutral theory, independently developed in 1968 by him and by Jack King and Thomas Jukes [6]. These writers shocked the biological world by arguing that the bulk of molecular evolution is due to selectively neutral mutations driven by the mutation process rather than selection. I think it would please Mayr that the general idea - that the rate of evolution in the population is equal to the rate of mutation in a single individual - can be derived by simple reasoning using school mathematics. Yet, in order to apply the idea, we need to know how long a time period must be observed. This depends on how long it takes for a lucky new mutant to increase in frequency and completely replace its predecessors. That is not a simple problem and requires sophisticated theory. Kimura solved it using a diffusion model (see [6]). When selection and migration are taken into account, the theory is much more complicated.

One contribution of the neutral theory has been to provide a rationale for a molecular clock. Essentially, all our estimates of evolution rates depend on the assumption that the molecular changes used in constructing the clock are mutation-driven. The near constancy of average mutation

**Figure 2**

(a) R A Fisher (1890-1962), (b) Sewall Wright (1889-1988) Photograph reproduced with permission from The Capitol Times, and (c) J B S Haldane (1892-1964).

rates permits reasonably accurate time estimates. Fortunately, enough of the DNA does not have an obvious function and can reasonably be supposed to be evolving by neutral kinetics, or near enough so that the neutral theory can be used in practice. And the experimenter can choose genomic regions most likely to behave in a neutral manner.

A second important attribute of the neutral theory is that it supplies a natural null hypothesis for the study of selection. And yet another outgrowth of the neutral theory is the view that much of the molecular polymorphism in natural populations is effectively neutral. This is especially useful now that variation in the frequencies of single-nucleotide polymorphisms (SNPs) is easily observed.

The various measures that are used to quantify genetic variability are outgrowths of population genetics theory. One striking result of such theory is the realization that all of the worldwide human population is descended from Africa, and moreover from a small area within Africa. The evidence for this striking conclusion is that molecular variance is greater in African peoples than elsewhere. The molecular clock can be used as one measure of the time taken during various human migrations and, of course, *Homo sapiens* is not the only species that can be studied in this way.

Another outgrowth of population thinking is the 'selective sweep'. A new favorable allele arises by mutation, spreads through the population and becomes fixed at a rate that is determined mainly by how favorable it is. A consequence of this fixation is that neutral or weakly selected alleles linked to the locus are swept along with it. Because of this, there is a region on either side of the selected locus that is deficient in genetic variability. Such regions of reduced variability are footprints of a selective sweep in the past and, remarkably, provide evidence for events that occurred long ago and which can no longer be observed. Although the basic idea is simple and requires no mathematics, an assessment of how

much the variability is reduced and the linkage distance over which the reduced variability occurs depend on mathematical theory.

Mathematics and the computation of family relationships

An area of biology in which mathematics, and especially computers, have become absolutely essential is systematics, Ernst Mayr's own field. Formerly, assessing species relationships and building phylogenetic trees based mainly on morphological differences was a matter of intuition and judgment. Systematists often disagreed, sometimes violently. Then came the DNA revolution. A mammalian DNA sequence supplies billions of bits of information, thus for the first time providing an opportunity for a procedure independent of personal judgments [7]. In recent decades the methods have steadily improved. The preferred procedures, such as Fisher's maximum likelihood, required a great deal of computation, and for a while this meant that large phylogenies were out of computer range. This is no longer true. Computers are now much faster, so their speed is no longer a limitation. Standards have increased in another way, too. It is now *de rigueur* to do statistical tests of significance of the tree structure and parts thereof. Many of these involve permutation methods, which have the merit of requiring minimum assumptions. They are computation-intensive, but with modern computers this is no longer an impediment.

One striking example from such studies, which came as a complete surprise to classical systematists, is the close relationship of the elephant to the shrew. Another example is in primates. For many decades the relationship of chimpanzee, gorilla and man has been uncertain. Molecular analysis of DNA sequences, using the newly developed theory, has shown that our closest relatives are chimpanzees. Furthermore, that we and the chimpanzees are 99% identical at the DNA level came as a surprise to many. Equally surprisingly, we share some 90% of our DNA with mice, rabbits, dogs, horses and elephants. Yet this is no surprise to those acquainted with the neutral theory. These numbers are fully consistent with expectations based on mutation rates and the times involved. Finally, there is now help available in the form of computer programs that can work out phylogenies and display the information graphically (see [7]). These not only eliminate a lot of tedious work, but place advanced methods in the hands of relative novices.

Coalescence and speciation

Finally, there has been a major theoretical advance, coalescent theory [8]. Instead of looking forward in time, this method looks backward. Any two alleles or homologous

nucleotides are ultimately derived from a single one; that is, looking backward, they coalesce. This has been the subject of extensive theoretical work in recent years. One problem for which coalescent theory provided at least an approximate answer is the question of whether there was any mating between our ancestors and contemporary Neanderthals. Small amounts of admixture are not ruled out, but coalescent theory has shown that any substantial intermating is very unlikely, as discussed by John Wakeley [8]. My other examples have been relatively simple, but this one isn't, as is apparent from this discussion. It involves a great deal of algebra. Another example, also given by Wakeley, is evidence for a selective sweep in *Drosophila simulans* [8].

Until recently, mathematical theory had contributed little to the study of speciation. Mayr emphasized allopatric speciation and the prevailing model, due to Dobzhansky and Muller [9], prevailed. Recent mathematical studies [10] support it and favor the view that speciation genes correspond to normal genes, selected for their effects within the species. Furthermore, there is evidence that these genes evolve rapidly. Thus, hybrid incompatibility is a by-product of ordinary selection in geographically isolated populations. There is no evidence that random drift plays an important part [9], so Mayr's 'genetic revolution' and similar ideas have little support. Yet it is important to point out that, aside from this, Mayr has usually been right [3]. The field of mathematical studies of speciation is barely started; it will surely increase.

Looking to the future

I have given only a few examples of the part that mathematical theory has played in evolution studies. There are many more, but these, I hope, constitute a convincing sample of the importance of mathematics in population genetics and evolution. I do not intend to imply that all evolutionary study need be mathematical and theory-driven. Much exciting evolutionary biology is done in the Mayr non-mathematical tradition [3]. For example, 'evo-devo' studies, looking at changes in development during evolution, have produced exciting results while largely ignoring population genetics. Another non-mathematical example is horizontal gene transfer brought about by transposable elements, which is especially important in the evolution of microorganisms. There is also abundant evidence for increases in genetic complexity by the accumulation of small duplications. And, as always, a lot of morphological and behavioral evolution is interesting in and of itself. Yet, my guess is that as these subjects become more quantitative, population genetic theory will play an increasing role.

The rise of molecular methods has led to an increase in the importance of mathematics in population genetics and evolution. The abundance of data that require mathematical analysis has greatly increased. At the time of Mayr's challenge, evolution had a beautiful theory but very few opportunities to apply it. Now the situation is reversed: data appear faster than existing theory can deal with them. That mathematics will play an increasingly important evolutionary role in the near future seems clear.

Envoi

I think these examples show not only that mathematical theory is helpful, but that it is often essential. I don't know what Ernst would say today. He might have had a change of mind, but I doubt it. Knowing how much he enjoyed arguing, I suspect he would be quite critical of much that I have written. Unfortunately, although he lived to be 100, he was not immortal and died in 2005. Were he still alive, I would surely hear from him and whatever his opinions, he would not keep them to himself. He would have enjoyed an argument, preferably over a glass of sherry. And so would I.

Acknowledgement

I am indebted to Bret Payseur for reading the manuscript and offering some very useful suggestions.

References

1. Mayr E: **Where are we?** *Cold Spring Harbor Symp Quant Biol* 1959, **24**:1-14.
2. Haldane JBS: **A defence of beanbag genetics.** *Perspect Biol Med* 1964, **7**:343-359.
3. Mayr E: *Populations, Species, and Evolution*. Cambridge, MA: Harvard University Press; 1970.
4. Crow JF: **Felix Bernstein and the first human marker locus.** *Genetics* 1993, **133**:4-7.
5. Barton NH, Charlesworth B: **Genetic revolutions, founder effects, and speciation.** *Annu Rev Ecol Syst* 1984, **15**:133-164.
6. Kimura M: *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press; 1983.
7. Felsenstein J: *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates; 2004.
8. Wakeley J: *Coalescent Theory*. Greenwood Village, CO: Roberts and Company; 2009.
9. Coyne JA, Orr HA: *Speciation*. Sunderland, MA: Sinauer Associates; 2004.
10. Orr HA: **The genetic basis of reproductive isolation: insights from *Drosophila*.** *Proc Natl Acad Sci USA* 2005, **102**:6522-6526.