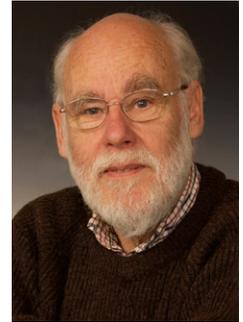


## What Use Is Population Genetics?

**Brian Charlesworth**

Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom



THE Genetic Society of America's Thomas Hunt Morgan Medal is awarded to an individual GSA member for lifetime achievement in the field of genetics. For over 40 years, 2015 recipient Brian Charlesworth has been a leader in both theoretical and empirical evolutionary genetics, making substantial contributions to our understanding of how evolution acts on genetic variation. Some of the areas in which Charlesworth's research has been most influential are the evolution of sex chromosomes, transposable elements, deleterious mutations, sexual reproduction, and life history. He also developed the influential theory of background selection, whereby the recurrent elimination of deleterious mutations reduces variation at linked sites, providing a general explanation for the correlation between recombination rate and genetic variation.

I am grateful to the Genetics Society of America for honoring me with the Thomas Hunt Morgan Medal and for inviting me to contribute this essay. I have spent nearly 50 years doing research in population genetics. This branch of genetics uses knowledge of the rules of inheritance to predict how the genetic composition of a population will change under the forces of evolution and compares the predictions to relevant data. As our knowledge of how genomes are organized and function has increased, so has the range of problems confronted by population geneticists. We are, however, a relatively small part of the genetics community, and sometimes it seems that our field is regarded as less important than those branches of genetics concerned with the properties of cells and individual organisms.

I will take this opportunity to explain why I believe that population genetics is useful to a broad range of biologists. The fundamental importance of population genetics is the basic insights it provides into the mechanisms of evolution, some of which are far from intuitively obvious. Many of these insights came from the work of the first generation of population geneticists, notably Fisher, Haldane, and Wright. Their mathematical models showed that, contrary to what was believed by the majority of biologists in the 1920s, natural selection operating on Mendelian variation can cause evolutionary change at rates sufficient to explain historical patterns of evolution. This led to the modern synthesis of evolution (Provine 1971). No one can claim to understand how evolution works without some basic under-

standing of classical population genetics; those who do run the risk of making mistakes such as asserting that rapid evolutionary change is most likely to occur in small founder populations (Mayr 1954).

---

**As our knowledge of how genomes are organized and function has increased, so has the range of problems confronted by population geneticists. We are, however, a relatively small part of the genetics community, and sometimes it seems that our field is regarded as less important than those branches of genetics concerned with the properties of cells and individual organisms.**

—B.C.

---

The modern synthesis is getting on for 80 years old, so this argument will probably not convince skeptical molecular geneticists that population genetics has a lot to offer the modern biologist. I provide two examples of the useful role that population genetic studies can play. First, one of the most notable discoveries of the past 40 years was the finding that the genomes of most species contain families of transposable elements (TEs) with the capacity to make new copies that insert elsewhere in the genome (Shapiro 1983). This led to two schools of thought about why they are present in the genome. One claimed that TEs are maintained because they confer benefits on the host by producing adaptively useful mutations (Syvanen 1984); the other believed that they are parasites,

maintained by their ability to replicate within the genome despite potentially deleterious fitness effects of TE insertions (Doolittle and Sapienza 1980; Orgel and Crick 1980).

The second hypothesis can be tested by comparing population genetic predictions with the results of TE surveys within populations. In the early 1980s, Chuck Langley, myself and several collaborators tried to do just this, using populations of *Drosophila melanogaster* (Charlesworth and Langley 1989). The models predicted that most *Drosophila* TEs should be found at low population frequencies at their insertion sites. This is so because *D. melanogaster* populations have large effective sizes ( $N_e$ ).  $N_e$  is essentially the number of individuals that genetically contribute to the next generation. Large  $N_e$  means that a very small selection pressure can keep deleterious elements at low frequencies. This is a consequence of one of the most important findings of classical population genetics—the fate of a variant in a population is the product of  $N_e$  and the strength of selection (Fisher 1930; Kimura 1962). If, for example,  $N_e$  is 1000, a mutation that reduces fitness relative to wild type by 0.001 will be eliminated from the population with near certainty.

Using the crude tools then available (restriction mapping of cloned genomic regions and *in situ* hybridization of labeled TE probes to polytene chromosomes), we found that nearly all TEs are indeed present at low frequencies in the population (Charlesworth and Langley 1989). Most of the exceptions to this rule were found in genomic regions in which little crossing over occurs (Maside *et al.* 2005). This is consistent with Chuck's proposal that a major contributor to the removal of TEs from the population is selection against aneuploid progeny created by crossing over among homologous TEs at different locations in the genome (Langley *et al.* 1988). It is now a familiar finding that nonrecombining genomes or genomic regions tend to be full of TEs and other kinds of repetitive sequences; the population genetic reasons for this, discussed by Charlesworth *et al.* (1994), are perhaps not so familiar.

Modern genomic methods provide much more powerful means for identifying TE insertions. Recent population surveys using these methods have confirmed the older findings: most TEs in *Drosophila* are present at low frequencies, and there is statistical evidence for selection against insertions (Barron *et al.* 2014). This is consistent with the existence of elaborate molecular mechanisms for repressing TE activity, such as the Piwi-interacting RNA (piRNA) pathway of animals (Senti and Brennecke 2010); there would be no reason to evolve such mechanisms if TEs were harmless. In a few cases, TEs have swept to high frequencies or fixation, and there is convincing evidence that at least some of these events are associated with increased fitness caused by the TE insertions themselves (Barron *et al.* 2014). These cases do not contradict the intragenomic parasite hypothesis for the maintenance of TEs; favorable mutations induced by TEs are too rare to outweigh the elimination of deleterious

insertions unless new insertions continually replace those that are lost.

---

**From the theory of aging, to the degeneration of Y chromosomes, to the dynamics of transposable elements, our understanding of the genetic basis of evolution is deeper and richer as a result of Charlesworth's many contributions to the field.**

—Charles Langley, University of California, Davis

---

My other example is a population genetics discovery about a fundamental biological process: the PRDM9 protein involved in establishing recombination hot spots in humans. This was enabled by the revolution in population genetics brought about by coalescence theory (Hudson 1990), which is a powerful tool for looking at the statistical properties of a sample from a population under the hypothesis of selective neutrality. The basic idea is simple: if we sample two homologous, nonrecombining haploid genomes (*e.g.*, mitochondrial DNA) from a large population, there is a probability of  $1/(2N_e)$  that they are derived from the same parental genome in the preceding generation; *i.e.*, they coalesce ( $N_e$  is the effective population size for the genome region in question). If they fail to coalesce in that generation, there is a probability of  $1/(2N_e)$  that they coalesce one generation further back, and so on. If  $n$  genomes are sampled, there is a bifurcating tree connecting them back to their common ancestor. The size and shape of this tree are highly random, so genetically independent components of the genome experience different trees, even if they share the same  $N_e$ . The properties of sequence variability in the sample can be modeled by throwing mutations at random onto the tree (Hudson 1990).

Recombination causes different sites in the genome to experience different trees, but closely linked sites have much more similar trees than independent sites. At the level of sequence variability, close linkage results in nonrandom associations between neutral variants—linkage disequilibrium (LD). The extent of LD among neutral variants at different sites is determined by the product of  $N_e$  and the frequency of recombination between them  $c$  (Ohta and Kimura 1971; McVean 2002). Richard Hudson proposed a statistical method for estimating  $N_e c$  from data on variants at multiple sites across the genome (Hudson 2001) that was implemented in a widely used computer program *LDhat* by Gil McVean and colleagues (McVean *et al.* 2002). Applications to large data sets on human sequence variability showed that the genome is full of recombination hot spots and cold spots, consistent with previous molecular genetic studies of specific loci (Myers *et al.* 2005). Most recombination occurs in hot spots and very little in between them, accounting for the fact that there is almost complete LD over tens or even hundreds of kilobases in humans. The identification of a large number of hot spots led to the discovery of

a sequence motif bound by a zinc finger protein, PRDM9, at about the same time that mouse geneticists also discovered that PRDM9 promotes recombination (McVean and Myers 2010; Baudat *et al.* 2014). These discoveries have led to many interesting observations, such as associations between PRDM9 variants in humans and individual variation in recombination rates, generating an ongoing research program of great scientific interest (Baudat *et al.* 2014).

With the ever-increasing use of genomic data, I am confident that many more such fruitful interactions between molecular and population genetics will take place. A take-home message is that more needs to be done to integrate training in population, molecular, and computational approaches to provide the next generation of researchers with the broad range of knowledge they will need.

### Literature Cited

- Barron, M. G., A. Fiston-Lavier, D. A. Petrov, and J. González, 2014 Population genetics of transposable elements in *Drosophila*. *Annu. Rev. Genet.* 48: 561–581.
- Baudat, F., Y. Imai, and B. De Massy, 2014 Meiotic recombination in mammals: localization and regulation. *Nat. Rev. Genet.* 14: 794–806.
- Charlesworth, B., and C. H. Langley, 1989 The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* 23: 251–287.
- Charlesworth, B., P. Sniegowski, and W. Stephan 1994 The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371: 215–220.
- Doolittle, W. F., and C. Sapienza, 1980 Selfish genes, the phenotype paradigm, and genome evolution. *Nature* 284: 601–603.
- Fisher, R. A., 1930 The distribution of gene ratios for rare mutations. *Proc. R. Soc. Edinb.* 50: 205–220.
- Hudson, R. R., 1990 Gene genealogies and the coalescent process. *Oxf Surv Evol Biol* 7: 1–45.
- Hudson, R. R., 2001 Two-locus sampling distributions and their application. *Genetics* 159: 1805–1817.
- Kimura, M., 1962 On the probability of fixation of a mutant gene in a population. *Genetics* 47: 713–719.
- Langley, C. H., E. A. Montgomery, R. H. Hudson, N. L. Kaplan, and B. Charlesworth, 1988 On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res.* 52: 223–235.
- Maside, X., S. Assimakopoulos, and B. Charlesworth, 2005 Fixations of transposable elements in the *D. melanogaster* genome. *Genet. Res.* 85: 195–203.
- Mayr, E., 1954 Change of genetic environment and speciation, pp. 157–180 in *Evolution as a Process*, edited by Huxley, J. S., A. C. Hardy and E. B. Ford. Allen and Unwin, London.
- McVean, G. A. T., 2002 A genealogical interpretation of linkage disequilibrium. *Genetics* 162: 987–991.
- McVean, G., P. Awadalla, and P. Fearnhead, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160: 1231–1241.
- McVean, G., and S. Myers, 2010 PRDM9 hits the spot. *Nat. Genet.* 42: 821–822.
- Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310: 312–324.
- Ohta, T., and M. Kimura, 1971 Linkage disequilibrium between two segregating nucleotide sites under steady flux of mutations in a finite population. *Genetics* 68: 571–580.
- Orgel, L. E., and F. H. C. Crick, 1980 Selfish DNA: the ultimate parasite. *Nature* 284: 604–607.
- Provine, W. B., 1971 *The Origins of Theoretical Population Genetics*. University of Chicago Press, Chicago.
- Senti, K.-A., and J. Brennecke, 2010 The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet.* 26: 499–509.
- Shapiro, J. A. (Editor), 1983 *Mobile Genetic Elements*. Academic Press, New York.
- Syvanen, M., 1984 The evolutionary implication of mobile genetic elements. *Annu. Rev. Genet.* 18: 271–293.