A century of variance

Brian Charlesworth and **Anthony W. F. Edwards** mark the 100th anniversary of a paper by R. A. Fisher, which introduced the statistical term "variance". What followed was a whole new field of statistical analysis

This year is the centennial of Sir Ronald Fisher's seminal paper, "The correlation between relatives on the supposition of Mendelian inheritance". Published in the *Transactions of the Royal Society of Edinburgh* on 1 October 1918,[1] the paper is notable for introducing the statistical term "variance" for the mean of the squares of the deviations of a measurement from its mean.

The analysis of variance, invented by Fisher, is one of the most powerful and widely used techniques in statistics. The basic idea is to break the total variance in a quantity of interest into different components associated with independent causal factors, together with a residual "error" variance. As later developed by Fisher, the *F*-ratio test can then be used to assess the statistical significance of the different causal factors, by comparison of their associated variances with the error variance.

The term "analysis of variance" does not appear in the body of Fisher's 1918 paper, but is used in the heading of Section 21 in the Contents. The paper does, however, deal with the partitioning of natural variability in biological traits into different causal components. A whole field of statistical analysis is opening up before our very eyes, which allows us to ask what proportion of the variance in a biological trait such as human height is determined by genetic differences among individuals, and what proportion is caused by non-genetic factors such as the effects of differences in nutrition. For example, modern research based on Fisher's ideas suggests that approximately 80% of the variance in human height among individuals within a single population is due to genetic factors.[2]

Fisher's paper was primarily concerned with a major biological question – the nature of the inheritance of quantitative traits in biology, such as human height or grain yield in crops, which show continuous or nearly continuous variability. Such traits are immensely important for the study of evolution, and for animal and plant breeding. Many human genetic diseases, such as high blood pressure, are of this type. Identifying the genetic variants underlying quantitative variation is now a major area of biological research,[3] and the conceptual framework introduced by Fisher is key to this enterprise.

Fisher and Darwin

Fisher's interest in the genetics of quantitative traits undoubtedly stemmed from his concern with evolution and admiration for the work of Charles Darwin, as well as his interest in eugenics.[4] Fisher's 1930 book, *The Genetical Theory of Natural Selection*, is widely regarded as the most original book on evolution after Darwin's *On the Origin of Species.* Indeed, Fisher's discussion of Darwin's ideas on variability in the first chapter of *The Genetical Theory* parallels the

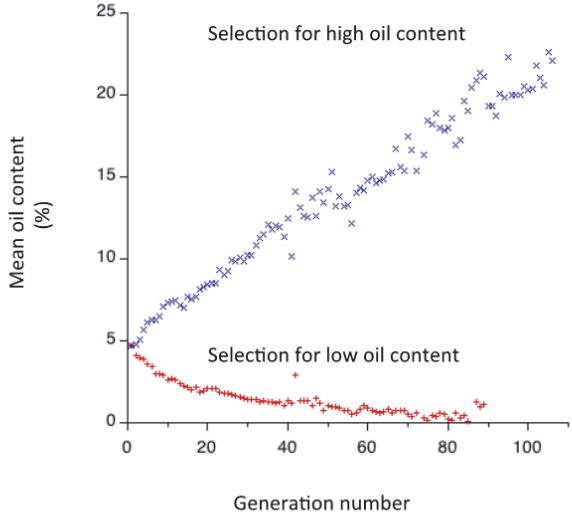

FIGURE 1 The world's longest-running experiment on artificial selection – selection for increased and decreased proportions of oil in the kernels of maize plants, initiated in 1896. Maize has one generation a year, so that 22 generations had been completed at the time Fisher's paper was published. In every generation, the top 20% of plants (high selection line) or bottom 20% (low selection line) were selected as parents. The blue crosses show the means for the high selection line and the red crosses the means for the low selection line. The mean for the final generation of the high line is more than 20 standard deviations greater than the initial mean, implying that there is little or no overlap between the two populations. The data were obtained from the Illinois long-term selection experiment for oil and protein content in corn, University of Illinois at Urbana-Champaign (hdl.handle.net/2142/3526).

LEFT Portrait of the English statistician and geneticist, Sir Ronald Aylmer Fisher (1890–1962). Fisher graduated in mathematics and physics from Cambridge University in 1912. He was appointed statistician at Rothamsted Experimental Station, in charge of 66 years of data on agricultural field trials. Fisher's contributions to statistics include methods of experimental design and maximum likelihood. In *The Genetical Theory of Natural Selection*, he showed that Mendel's genetic laws and Darwin's theory of natural selection are in full accord. Fisher became professor of genetics at Cambridge in 1943, and was knighted in 1952. Credit: Science Photo Library

way in which Darwin started *On the Origin of Species* with a discussion of variability in domesticated animals and plants.

In his first chapter, Darwin showed that selective breeding by humans has produced dramatic changes in the characteristics of domestic animals and plants of agricultural importance. His aim was to demonstrate that evolutionary change in natural populations could be brought about by natural selection acting on variability of the same kind that had been exploited by animal and plant breeders. If individuals vary with respect to a particular trait (e.g. the leg lengths of a deer species), and the trait affects the survival or reproductive success of individuals (deer with longer legs can run faster than deer with shorter legs, thereby escaping from wolves more easily), the survivors will differ in mean from the overall population. If the trait has a genetic basis, so that offspring resemble their parents, the mean of the offspring generation will differ from that in the previous generation. If this process is continued generation after generation, a large change in the mean eventually results. Indeed, carefully controlled selection experiments have shown that a relatively small number of generations of strong selection can produce individuals with trait values that are outside the limits seen in the initial population (Figure 1, page 21).

While Darwin's evidence was sound, and the conclusions that he drew are broadly correct in the light of modern knowledge, he was hampered by a lack of understanding of inheritance, as described by Fisher in *The Genetical Theory*. Darwin subscribed to the belief that inheritance was "blending", so that the offspring of a mating between two individuals would have a genetic make-up that was halfway between the two parents. As was pointed out in 1864 by the Edinburgh professor of engineering, Fleeming Jenkin, this has disastrous consequences for genetic variability, which rapidly disappears from the population. If mating is random with respect to the trait values of the parents, the genetic component of variance is halved every generation. Since responses to selection require genetic variability, this is inimical to the plausibility of natural selection. Darwin's response was to appeal to the replenishment of variability by direct effects of the environment on the traits of parents, which he postulated could be transmitted to the offspring. While there is no doubt that the environment can affect many traits, by the time Fisher was writing in 1930 experimental genetics had discredited the idea that such environmental effects are often transmissible across generations.

Darwin was not himself a quantitative thinker, but was in close contact with his first cousin, Francis Galton. Galton was one of the first to collect data on the resemblances between relatives, especially human height. For this purpose, Galton introduced the terms "regression" and "correlation" in 1886, but without using the standard "least squares" method for estimating regression and correlation coefficients, which emerged from later work by Edgeworth, Pearson and Yule. In the early 1900s, Pearson and his collaborators used these methods to conduct extensive analyses of the resemblances between relatives for quantitative traits.[5] For human height, the correlation coefficient (which takes a value of 0 for no relationship and 1 for a perfect relationship) between a parent and an offspring was approximately 0.45, about half this value for grandparents and grandchildren, 0.54 for full siblings, and so on. Fisher employed these values in his 1918 paper. At the time he was writing, there was essentially no dispute about the correctness of the biometrical estimates, but there were violent disagreements about how to interpret them and about their biological meaning.

About the authors

Brian Charlesworth is an evolutionary biologist and population geneticist in the Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh.

Anthony W. F. Edwards is a statistician, geneticist, and evolutionary biologist, and a fellow of Gonville and Caius College, Cambridge.

Biometry meets Mendelism

To understand the source of these controversies, it is necessary to examine the history of genetics in the early twentieth century. Early experimental genetics, which effectively started in 1900 with the rediscovery of the 1866 paper of Gregor Mendel, concentrated on easily classified discrete (or "discontinuous") traits, such as the round versus wrinkled peas that Mendel studied. It was found that these alternatives were controlled by different forms (*alleles*) of the same factor (*gene*), and that no blending takes place when two alternative alleles are present in the same individual (Figure 2).

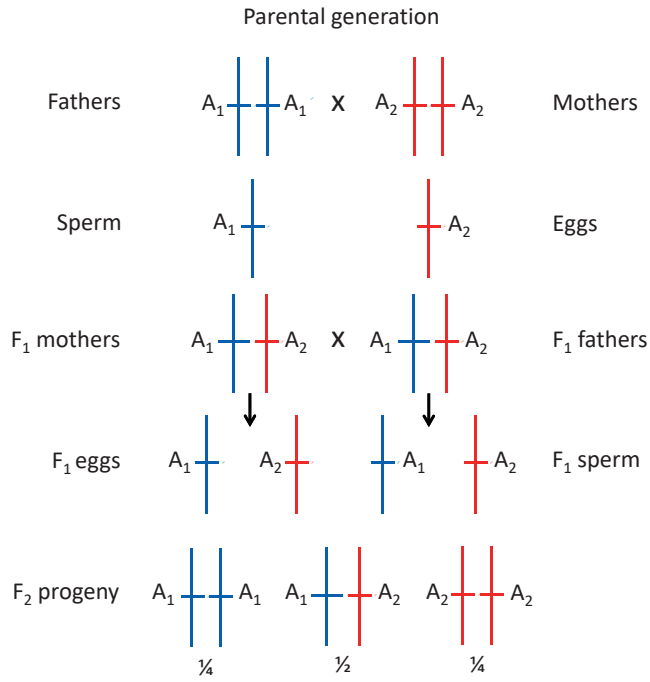

FIGURE 2 The core principles of Mendelian genetics. The blue and red lines indicate the DNA molecules of which the gene in question is composed; each adult individual has two copies, one of maternal and the other of paternal origin. During the production of eggs and sperm, a specialised type of cell division ensures that either the maternal or the paternal copy is passed at random into the egg or sperm cell. The fusion of an egg and a sperm re-establishes the state of the adult. The horizontal lines indicate the locations of two variant forms of the DNA sequence of the gene, denoted by A_1 and A_2. The figure shows what happens when two parents are crossed to form an initial F_1 generation, and their progeny are crossed among themselves to produce a second F_2 generation.

Note that each parental type (A_1A_1 and A_2A_2) is produced in the F_2 generation with a probability of one-quarter, and the F_1 type (A_1A_2) with a probability of one-half. In a cross where the two parents differed with respect to n independent genes, the probability of obtaining a given parental type in the F_2 generation would be $(1/4)^n$; with 10 genes, this is only one in a million.

We now know that such allelic differences correspond to differences in the DNA sequence of the gene whose product is involved in the development of the trait in question.

The early Mendelian geneticists, led by the outspoken William Bateson, thought that such discontinuous traits were the raw material for evolution and downplayed the importance of continuous variability of the type shown in Figure 1. If natural selection played a role at all, it simply selected the occasional beneficial *mutation* (a sudden change in a gene), causing the mutation to spread through the population. Galton himself had toyed with the idea of such a discrete basis for heredity, but Pearson vehemently rejected it, and emphasised the importance of continuous, quantitative variation. This led, in Britain at least, to a sustained battle between the biometricians and the Mendelians.[5]

There is, however, no contradiction between discrete inheritance at the level of the genetic material itself and apparently continuous variability at the level of traits such as human height and the oil content of maize kernels described in Figure 1. Fisher's 1918 paper was the most profound and lucid description of how to reconcile biometry with Mendelism. It is worth pointing out, however, that the explanation – often known as the multiple factor theory of quantitative trait inheritance – had already been suggested by Mendel himself, and by 1910 had been empirically tested by two plant geneticists, Herman Nilsson-Ehle in Sweden and Edward East in the USA.[5]

Figure 3 shows the basic principle involved: the trait is influenced by several different genes, whose effects in this example combine additively. There is also variability arising from non-genetic causes. If the individual effects of the genes are sufficiently small in relation to the overall level of variability in the trait, and the number of genes affecting the trait is sufficiently large, a continuous range of variability will result. In contradiction to the expectation under blending inheritance, the variability in the final generation after a cross between two stocks will slightly exceed that in the second generation. If you raise enough individuals, that is exactly what is found.

This is fine for understanding the results of experimental crosses, but does not help with the interpretation of the biometrical results on correlations between relatives in freely breeding populations, as Pearson pointed out. To achieve such an interpretation, we need mathematical models of how the effects of naturally occurring variants in individual genes produce variability at the level of the traits they control, and generate the observed patterns of correlations among relatives. The construction and analysis of such models was the major contribution of Fisher's 1918 paper.

Fisher's treatment

A fundamental first step towards achieving this goal was the realisation that Mendelian inheritance implies that variability is maintained in a population if no evolutionary forces are acting. If we have two different allelic forms of a gene, which we can call A_1 and A_2, the state of the population with respect to this gene can be described by the frequencies *P*, *Q* and *R* of

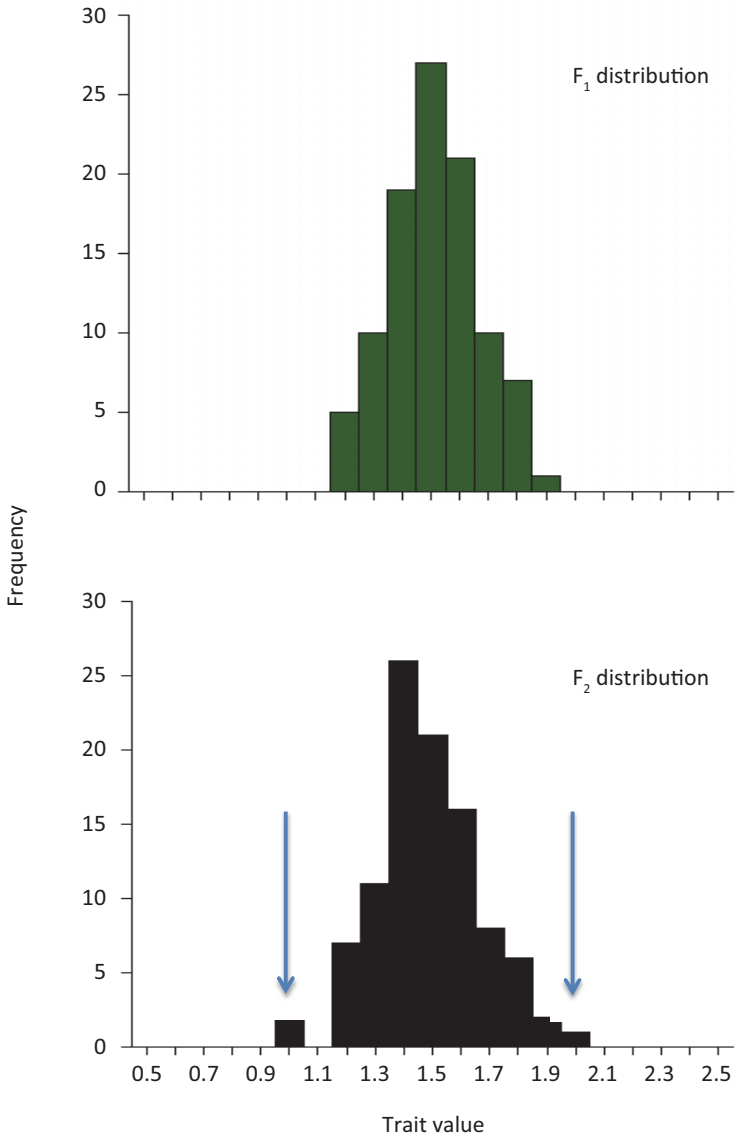

FIGURE 3 The joint effects of non-genetic variability and of multiple genes with small effects, modelled by a computer simulation. Here, two parental lines that differed with respect to 10 independent genes affecting a quantitative trait were crossed to produce 100 progeny in the F_1 generation. The genes had completely additive effects on the trait. The F_1 individuals were completely genetically uniform, so that the green histogram shows the effect of purely environmental sources of variability. As indicated by the blue arrows, the means of the parents were 1 and 2. The F_1 mean was approximately 1.5, with an estimated variance of 0.0238 (the true variance for the distribution of environmental effects was 0.025). The F_1 progeny were crossed together to produce 100 individuals in the next generation (F_2), shown by the black histogram. Their mean was also approximately 1.5, and their variance was 0.0335, slightly larger than the F_1 value.

Use of Fisher's *F*-ratio test to compare the two variances gave a probability of just under 0.05 that such a large difference could occur by chance if the true variances were the same (the *p*-value). This would normally be regarded as only marginally significant. Re-running the simulations to generate 1000 F_2 progeny gave a ratio of F_2 to F_1 variances of 1.48, which corresponds to a *p*-value of 0.01, indicating statistical significance.

This simulation shows why biologists believed for a long time that inheritance was blending – it is difficult to detect an increase in variance in the F_2 generation without large samples, and crosses between lines are often approximately intermediate between the parents. It also illustrates the need for proper statistical procedures when conducting tests of genetic hypotheses; it is no accident that Fisher made path-breaking contributions to both statistics and genetics.

the three possible combinations (*genotypes*) that compose the population: A_1A_1, A_1A_2, A_2A_2. If there is no selection, and the population is so large that random fluctuations in the frequencies are negligible, the frequencies will remain constant from generation to generation. If individuals mate randomly, the frequencies of the three genotypes are p^2, $2pq$ and q^2, where $p = P + \frac{1}{2}Q$ and $q = R + \frac{1}{2}Q$ are the frequencies of the alleles A_1 and A_2, respectively. This is the famous *Hardy–Weinberg law*, formulated independently by the English mathematician G. H. Hardy and the German physician Wilhelm Weinberg in 1908.[5] As Fisher eloquently described in the first chapter of *The Genetical Theory*, this result resolves Darwin's difficulty with loss of variability: the mechanism of heredity preserves rather than destroys variability.

The next step is to associate the different genetic types with different trait values, and to use the resulting model to calculate the quantities of interest. It is important to note that Fisher was not the first person to do this.[5] In fact, in 1904 Pearson had produced a rather restricted model, which he claimed produced results that were incompatible with the biometrical results. Yule pointed out, however, that relaxing the assumptions made by Pearson would remove the apparent inconsistency. In 1910 Weinberg went much further and gave a detailed mathematical treatment of the problem, which in many ways is similar to Fisher's; his work has been unjustly neglected.[5] Fisher's work was very much a response to Pearson, but he was unaware of Weinberg's papers. Similarly, Sewall Wright in the USA developed another independent treatment, but using a more restricted model than either Fisher or Weinberg.[5]

Fisher's treatment has therefore been the most influential of this early work, and displays both his characteristic insight into the biological problem and his mathematical ingenuity. The basic model is extremely simple. If we take a particular gene, we can imagine that we can measure the trait value for individuals of types A_1A_1, A_1A_2, and A_2A_2, averaged over the whole population. Let these *genotypic values* be represented as $-a$, d, and a, respectively. If $d = 0$, the two different forms of the gene combine additively to produce the trait value of the A_1A_2 individuals; if $d \neq 0$, then there is some degree of *dominance* of one type over the other. Fisher showed that the total variance contributed by a gene could be split into two components, one reflecting the variance explained by the linear regression of the trait value of a genotype on the number of copies of the A_2 allele that it contains (0, 1, or 2), and the other reflecting the remaining component. These are now referred to as the *additive* and *dominance* variances due to the gene in question. For this single gene case, with frequencies p and q of alleles A_1 and A_2, the additive variance (V_A) and dominance variance (V_D) in a randomly mating population are $V_A = 2pq[a + d(p - q)]^2$ and $V_D = (2pqd)^2$. In the absence of dominance, all the genetic variance is additive.

Fisher showed that this approach can be extended to the effects of many genes by assuming that the trait value of a particular individual can be represented by the sum of the genotypic values of each gene it contains, together with a term that represents the deviation from additivity of the effects of different genes, called *epistasis*. The total variance in the trait due to genetic causes is the sum of the additive, dominance and epistatic variances, plus a term representing the contribution of non-genetic effects and their interactions with the genetic effects.

He went on to show how the correlations between relatives could be related to the genetic variances, by deriving algebraic formulae based on writing out the genotypes produced by the relevant matings. In practice, he mostly ignored the epistatic variance, but allowed for dominance. Remarkably, the correlations among relatives then depend only on the additive and dominance variances, and not on the underlying genotype frequencies. He also considered the difficult problem arising from the fact that humans do not mate randomly with respect to quantitative traits such as height; there is a strong positive correlation between partners with respect to many traits (the correlation for height between father and mother is 0.28), and this causes the correlations between relatives to be higher than with random mating. This is the most intricate part of the paper, which has stood the test of time.

What were his conclusions? Table 1 shows examples of the expected correlations among different types of relatives, for the case of a randomly mating population, using modern notation and ignoring epistatic variance effects (work in the 1950s by Clark Cockerham and Oscar Kempthorne gave explicit formulae for the epistatic contributions, which need to be broken down into subcomponents for this purpose). The patterns of expected correlations closely parallel those found empirically by Galton and Pearson, and in many subsequent studies, showing that the biometrical data are consistent with Mendelian inheritance. These correlations can be used to estimate the contributions of the individual variance components. For example, with a correlation coefficient for height of 0.45 between parent and offspring, the relevant formula in Table 1 implies that 90% of the variance in height is additive genetic in origin.

Fisher was particularly interested in the finding that the correlation between parent and offspring involves only the

TABLE 1 The expected correlations between different types of relatives arising from genetic causes.

Type of relative	Expected correlation
Identical twins	V_G/V_T
Parent–child	$0.5V_A/V_T$
Full siblings	$(0.5V_A + 0.25V_D)/V_T$
Grandparent–grandchild	$0.25V_A/V_T$
Uncle (aunt)–nephew(niece)	$0.25V_A/V_T$
Half siblings	$0.25V_A/V_T$
Double first cousins	$(0.25V_A + 0.0625V_D)/V_T$
Great-grandparent–great-grandchild	$0.125V_A/V_T$
Single first cousins	$0.125V_A/V_T$

V_A is the additive genetic variance; V_D is the dominance variance; V_G is the total genetic variance; V_T is the total variance in the trait. Epistatic contributions to the total genetic variance and environmental causes of correlations between relatives have been ignored.

additive variance, while the correlation between full siblings (0.54) includes a contribution from the dominance variance. He called the ratio $V_D / (V_A + V_D)$ the "dominance ratio", and estimated its value as approximately 0.3 from data of Pearson and Lee on human body measurements. Later work has shown that correlations between siblings due to the effects of shared environments are probably the major contributor to the excess correlation among siblings, and that additive variance is the major genetic component for most quantitative traits.[2] Modern genetic studies have, however, validated Fisher's basic postulate that genetic variation in traits such as height is caused by numerous independent genes with very small individual effects.[3]

A whole branch of genetics flowed from Fisher's 1918 paper, which also laid the foundations for his later ground-breaking statistical work

A lasting influence

Fisher's 1918 paper ends with a section, "The Interpretation of the Statistical Effects of Dominance", which acts as a sort of trailer for the discussion of selection in his important 1922 paper,[6] to which much of modern population genetics can be traced. The structure of the methods of analysis advanced had a profound, and largely unnoticed, influence on the development of the theory of natural selection in Fisher's 1930 book. As described earlier, Darwin argued that genetic variation in a character, coupled with the forces of natural selection, allows change in the state of the character in subsequent generations, causing it to evolve. Mendel provided an understanding of the mechanism of inheritance, enabling the early population genetic models of the Darwinian process to be set up. Fisher's approach was different. He studied the mathematics of the variance of the character itself, how it was maintained and what influences bore upon it. He sought a quantitative version of Darwin's theory: how fast would the mean of a character change, given its current variance?

In 1930, Fisher adopted his 1918 point of view, defining the additive genetic variance in terms of the regression described above. He proved that, if the character in question is genotypic fitness itself, then the contribution to the rate of change of its mean arising from changes in allele frequency is equal to the additive genetic variance; one way of looking at this is to note that it is the additive genetic variance alone that determines the resemblance between parent and offspring (see Table 1). A modern rewording of Fisher's 1930 original statement is:

"The rate of increase in the mean fitness of any organism at any time ascribable to natural selection acting through changes in allele frequencies is exactly equal to its additive genetic variance in fitness at that time."

This is Fisher's fundamental theorem of natural selection, the centrepiece of *The Genetical Theory of Natural Selection*. Darwin's theory had been quantified. Natural selection does not necessarily increase the mean fitness of a population. However, as it unceasingly alters the frequencies of alleles it contributes a defined and positive amount to the rate of change of the mean fitness, given by the additive genetic variance of fitness. We can also apply the theorem to a character that is correlated with fitness, replacing the variance with the covariance between the character and fitness. (Fisher introduced his argument by taking the character "human height" as an example, but finally developed it for fitness itself, pointing out that fitness is after all a genotypic character.) This version of the theorem has considerable practical value, since it allows animal and plant breeders to predict the speed of response of a character to selection from measurements of its additive genetic variance.

A whole branch of genetics of both theoretical and applied importance flowed from Fisher's 1918 paper, which also laid the foundations for his later ground-breaking statistical work on the analysis of variance and design of experiments. It is fitting that the Fisher Memorial Trust will be celebrating the centenary of this seminal work with a one-day scientific meeting in Edinburgh on 9 October 2018, in collaboration with the Royal Statistical Society, the London Mathematical Society, the Genetics Society, and the Galton Institute (bit.ly/fishercentenary). ■

Acknowledgements

We thank Deborah Charlesworth, Bill Hill and Peter Visscher for their helpful comments on the first draft of this article. Comments from two reviewers helped to improve it.

References

1. Fisher, R. A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, **52**, 399–433.

2. Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M. and Posthuma, D. (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, **47**, 702–709.

3. Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A. and Yang, J. (2017) 10 years of GWAS discovery: Biology, function, and translation. *American Journal of Human Genetics*, **101**, 5–22.

4. Norton, B. and Pearson, E. G. (1976) A note on the background to, and refereeing of, R.A. Fisher's 1918 paper 'On [sic] the correlation between relatives on the supposition of Mendelian inheritance'. *Notes and Records of the Royal Society of London*, **31**, 151–162.

5. Provine, W. B. (1971). *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press.

6. Fisher, R. A. (1922). On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, **42**, 321–341.