

Polymorphism

Oliver Mayo
*CSIRO Livestock Industries, Adelaide
Australia*

1. Introduction

"I refer to those genera which have sometimes been called 'protean' or 'polymorphic,' in which the species present an inordinate amount of variation; and hardly two naturalists can agree which forms to rank as species and which as varieties. ... I am inclined to suspect that we see in these polymorphic genera variations in points of structure which are of no service or disservice to the species, and which consequently have not been seized on and rendered definite by natural selection, as hereafter will be explained." (Darwin 1859 Ch. 2) Although Darwin was pointing to taxonomic problems caused by meaningless variation here, he clearly understood that a species could manifest variations that were neutral in the face of natural selection, and hence were not removed by natural selection. With no explicit demographic or genetical model, Darwin could not take the discussion further, but the concept of polymorphic variation within a species is clearly 150 years old at least.

Once genetics had been set on a sound footing by Mendel and his rediscoverers, genetic polymorphisms were rapidly identified. Sex determination was one of the first and most important; other outbreeding mechanisms, such as heteromorphic self-incompatibility, a major subject of Darwin's own research, were soon identified as functional polymorphisms.

Polymorphism was thus identified as variability that was genetically determined. How it related to other phenotypic variability was not clear. At the time when Mendelian genetics and statistically measured quantitative genetic variation were reconciled by Fisher (1918), the role of individual genes in influencing quantitative variation was barely initiated, through the study of, for example, dwarfing genes.

2. Balanced polymorphism

"If selection favours the homozygotes, no stable equilibrium will be possible, and selection will then tend to eliminate whichever gene is below its equilibrium proportion; such factors will therefore not commonly be found in nature: if, on the other hand, the selection favours the heterozygote, there is a condition of stable equilibrium, and the factor will continue in the stock. Such factors should therefore be commonly found, and may explain instances of hybrid vigour, and to some extent the deleterious effects sometimes brought about by inbreeding." (Fisher 1922 p. 324)

The argument, which introduces notation etc., is as follows. Consider a diallelic locus with two alleles, A_1 and A_2 , having frequencies p and q in an indefinitely large population with

random mating. Then the population will be in Hardy-Weinberg equilibrium, as is well known (see e.g. Mayo 2008 for review). Hardy-Weinberg equilibrium (HWE) is the state of the genotypic frequencies of two alleles of a single gene (locus) after one generation of random mating in this infinitely large population with discrete generations, in the absence of migration, mutation and selection: if the alleles are A_1 and A_2 with frequencies p and q ($= 1-p$), then the equilibrium gene frequencies are just p and q and the equilibrium genotypic frequencies for $A_1 A_1$, $A_1 A_2$ and $A_2 A_2$ are p^2 , $2pq$ and q^2 . Thus, there is equilibrium at both the allelic and the genotypic level. Table 1 gives the frequencies for the 3 genotypes revealed by gel electrophoresis of human red cell adenylate kinase in a number of European populations and one non-European population. (Here, q , the frequency of the rarer allele, is estimated as $\text{fr}(AK_2 AK_2) + \frac{1}{2}\text{fr}(AK_1 AK_2)$.) This table illustrates differences among populations arising from geography and other factors, so that such variation can be used for purposes such as investigating the history of population growth, migration etc.

Population	Genotype			Sample size	Frequency of AK_2
	$AK_1 AK_1$	$AK_1 AK_2$	$AK_2 AK_2$		
British	1720	165	2	1887	0.0448
Indian (England)	107	24	1	132	0.0985
Irish	739	50	0	789	0.0317
Finland	71	6	0	77	0.0390
Finnish Lapps	304	3	0	307	0.0049
Germany (Berlin)	1865	142	1	2008	0.0359
Germany (SW)	382	25	0	407	0.0307
Italy (Rome)	686	52	0	738	0.0352
Italy (Sardinia)	1004	28	1	1033	0.0145

Table 1. Genotype and gene frequencies for the red cell adenylate kinase locus in a number of samples from different human populations (extracted from Tills *et al.* 1971)

Most of the populations in Table 1 are in Hardy-Weinberg equilibrium, suggesting limited effects of disturbing factors, which include migration, inbreeding and natural selection. Now suppose that there is selection against all genotypes such that the fitnesses of the genotypes are as shown:

	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	Total
Fitness	α	β	γ	
Birth frequency	p^2	$2pq$	q^2	1
Freq. post-selection	$p^2\alpha/T$	$2pq\beta/T$	$q^2\gamma/T$	T

After selection, frequency (A_1) = p'

$$= (p^2\alpha + pq\beta)/T$$

The change in p ,

$$\begin{aligned} \Delta p &= p' - p \\ &= (p^2\alpha + pq\beta - pT)/T \end{aligned}$$

If selection is not to alter gene frequencies, i.e. there is a gene frequency equilibrium, then $\Delta p = 0$ i.e. $p^2\alpha + pq\beta = pT$, whence $p=0$ & $q=1$ or $p=1$ & $q=0$ or, non-trivially,

$$p = (\gamma - \beta) / (\alpha - \gamma - 2\beta) \text{ \& } q = (\alpha - \beta) / (\alpha - \gamma - 2\beta) .$$

This last solution requires that $\alpha < \beta$ and $\gamma < \beta$. Hence, we usually write $\alpha = 1-s$ $\beta = 1$ $\gamma = 1-t$.

Related results hold for an X-linked gene and multiple alleles at an autosomal locus (see Mayo 1978 and Bürger 2000). The results for autosomal loci can be very complex: 'even in the absence of epistasis, mean fitness is not necessarily increasing; linkage disequilibrium may persist forever; and completely polymorphic stable equilibria may coexist.' (Bürger 2000, p. 51) (Here, epistasis is any interaction between non-allelic genes, not simply the suppression of the effect of one gene by another non-allelic gene.)

Following Fisher's pioneering work, Ford (1940, 1964) redefined genetic polymorphism as 'the occurrence together in the same locality of two or more discontinuous forms of a species in such proportions that the rarest of these cannot be maintained merely by recurrent mutation.' Ford then emphasised that such polymorphism could be balanced, as in Fisher's theoretical argument above, or transient, whereby neutral or nearly neutral alleles of a gene had frequencies fluctuating by chance, and all but one would eventually be lost from the population. Without this contrast, Ford could have been accused of assuming what he wanted to discover, namely, balancing natural selection, given that no alleles would be lost from an infinitely large population (and little was known of natural population sizes at the time), and that the rate of loss of genetic variability from large populations had been shown by Fisher and Wright to be very slow (Fisher 1922, 1930, Wright 1930).

Ford and others searched for balanced polymorphisms, and found some cases. An important one, which influenced much thinking over a long period (Mayo 2007), is sickle-cell anaemia in humans. Here, sickle-cell homozygotes ($Hb\beta^S Hb\beta^S$) have defective haemoglobin and can suffer from pernicious anaemia $\gamma=0.25$, heterozygotes ($Hb\beta^A Hb\beta^S$) are resistant to malaria and have normal erythrocytes $\beta=1$, and normal homozygotes suffer badly from malaria $\gamma=0.8$. This polymorphism results from one gene affecting two traits, rather than epistasis, two genes affecting one phenotype, though there are several other genes that affect response to the malaria parasite. Given the wastage through illness and death, there cannot be many such polymorphisms in a population at any one time.

Wastage (or suffering) is necessary to remove deleterious mutations from a population: as Sved (2007, p. 461) put it, "one 'genetic death' is necessary to remove a deleterious mutation, no matter how small the effect of that mutation." These "genetic deaths", however manifested, constitute genetic load (see Morton 2007 for the history of the concept). If μ is the mutation rate to deleterious alleles at a locus, then the load is μ for recessive, 2μ for dominant mutations. One of the reasons for the idea of genetic load being discarded, except through acknowledgement of the 'burden felt in terms of death, sterility, illness, pain and frustration' as a result of deleterious mutation (Crow 1970) i.e. the human condition, as unfruitful has been the recognition of the sheer scale of variation at the DNA level, that is to say, polymorphism. For conclusive work on the theory of mutational load and how it relates to population variability, see Bürger and Hofbauer (1994) and Bürger (2000).

The concepts inherent in the stable equilibria discussed above are influenced by population size in three distinct ways. First, in a finite population there is a non-zero probability that

one of the two alleles will be lost by chance. Secondly, in a population of effective size (the number of randomly mating individuals that give the same population dynamics as the population in question: Bürger 2000) N_e , with a mutation rate to new alleles μ for a given gene, $1+4 N_e\mu$ alleles of that gene can be maintained (Kimura and Crow 1964). Equivalent results hold for X-linked loci (Mayo 1976, Nagylaki 1992). Thus, observation of polymorphism of a gene need have no implication in regard to selection. Thirdly, if $s > 5t$ or $t > 5s$ and $2N_e(s+t) < 8$, the polymorphism will actually be lost rapidly from the population (Robertson 1962, Ewens and Thomson 1970). Mayo (1971, p. 329) noted that, 'although tetraploidy enhances the conservation of variation in small populations, it does not appear to do so by a factor of 2, as might have been expected from a doubling of the amount of genetic material.'

3. Linkage equilibrium

Suppose that there are two genes, A with alleles A_1, A_2 having frequencies p_A, q_A and B with alleles B_1, B_2 having frequencies p_B, q_B . There are 4 possible gametes

	$A_1 B_1$	$A_1 B_2$	$A_2 B_1$	$A_2 B_2$
frequencies	x_1	x_2	x_3	x_4
equilibrium	$p_A p_B$	$p_A q_B$	$q_A p_B$	$q_A q_B$

At equilibrium,

$$x_1 x_4 = x_2 x_3$$

The departure from equilibrium, $D = x_1 x_4 - x_2 x_3$ is termed 'linkage disequilibrium' (LD). It is easy to show (see e.g. Bürger 2000) that if the recombination between the A and B loci is c and LD in generation t is D , then in generation $t+1$ it is $D' = (1-c)D$. If the genes are unlinked, $D' = \frac{1}{2}D$. If linkage is very tight, i.e. $c \ll 0.5$, D will decline very slowly.

The importance of observed values of D can readily be tested statistically, because on the hypothesis that $D = 0$, $nD^2 / (p_A p_B q_A q_B)$ will be distributed approximately as χ^2 with one degree of freedom. (Because it is the square of the correlation between the loci, $r^2 = d^2 / p_A(1-p_A) p_B(1-p_B)$ is frequently used as the variable of interest.) For more than two alleles, equivalent results have been presented by Zaykin *et al.* (2008). Although much more complex because of the volume of data to be analysed, the analysis of LD data rests on modest, well understood statistical tools.

From the discussion above, LD between closely related polymorphisms is expected to be the norm, if there are large numbers of polymorphisms, unless these arose in linkage equilibrium and all populations are very large. Now several million single nucleotide polymorphisms (SNPs; section 5) are known within human populations, which means that the mean distance between human SNPs is about 1000 base pairs, the genome having about 3,000,000,000 base pairs (International Human Genome Sequencing Consortium, 2004). In the important experimental animal *Drosophila melanogaster*, with a much smaller genome (<200,000,000 base pairs), a population may contain up to one million SNPs (Burke *et al.* 2010). Hence, it is to be expected that there will be many closely linked polymorphisms. To make this statement more strongly, note that if the average human chromosome is about 2.5 Morgans long in recombination terms, and contains over 100,000 polymorphisms, there will

be about 400 polymorphisms/1% recombination. Hence, linkage disequilibrium must be the norm, unless, as noted above, most polymorphisms arose in linkage equilibrium with each other. Even in this unlikely eventuality, for which there is no evidence, finite population size would generate linkage disequilibrium rapidly (Haldane 1940, Sved 1968).

Because of these stochastic effects on linked genetic variation, populations that are reproductively isolated from each other rapidly become genetically different, and this fact can be used to estimate time of divergence and other attributes on relatively simple assumptions. For example, Sved (2009) has shown that k small populations of effective size will develop r^2 equal to $1/[1+4N_e c(1+(k-1)\rho)]$, where $\rho = m/[c(k-1)+m]$, m being the rate of migration between adjacent populations. Applying this simple model to African and European HapMap (see 7.2 below) data, Sved (2009) obtained an estimate of time of divergence of less than 1000 generations, if there were relatively little migration subsequent to the original separation. This estimate is about one-third of current minimum estimates of divergence time, illustrating both how models can oversimplify complex situations and how much more is to be learnt about migration out of Africa.

Maynard Smith and Haigh (1974) first quantified the effect of selection of alleles of one gene on a linked gene. They showed that such selection would generate disequilibrium on a chromosome, the effect decaying with increasing distance from the gene under selection. This effect, termed 'hitchhiking', is important when selection is intense or rapid. As a selected allele increases in frequency in a population, or 'sweeps through' a population, adjacent genes show allele frequency changes, and even mildly deleterious alleles of neighbouring genes may rise in frequency. Overall, many changes will be wrought by a selective sweep, and their interpretation may not be straightforward. For example, Rose *et al.* (2011) showed that selection for organophosphorus insecticide resistance in the Australian sheep blowfly, *Lucilia cuprina*, changed allele frequency at the primary locus, α -esterase, whose alleles conferred resistance but also changed frequencies at many adjacent loci, including altering structural polymorphisms. This work supports the general concept that micro-evolution of this kind depends 'primarily on pre-existing intermediate-frequency genetic variants that are swept the remainder of the way to fixation' (Burke *et al.* 2010, p. 587). In fact, as Burke *et al.* further conclude, a so-called 'soft sweep' of this kind is unlikely to lead to fixation because of fluctuating selective forces and the likelihood that the selected alleles have pleiotropic effects that are not advantageous.

4. Sex determination and related polymorphisms

Inbreeding is generally deleterious. Self-fertilisation is the most extreme form of inbreeding. The requirement for two sexes in most multicellular organisms helps to ensure avoidance of selfing. In most animals, the two sexes are involved in the production of offspring. The human XY chromosome male XX chromosome female is one example of many; sex-determination can be by alleles of one gene, or by a number of different chromosomal differences.

In an indefinitely large population, the XX-XY system is stable and leads to equal frequencies of the two sexes (Fisher, 1930). In a small population, there is a low but non-zero possibility that an entire generation will be of one sex, so that the population dies out. Any given Y (or X) chromosome will eventually die out, and (because of the association of the Y

chromosome with family name in some human populations) this possibility was first studied and branching processes used, in studies of the disappearance of family names (Watson and Galton 1875).

All sex-determining systems have advantages and disadvantages. The XX-XY system, for example, allows the sheltering of lethal alleles in the heteromorphic segment of the Y (Muller 1932), and hence the steady loss of function from the Y: 'the genes of the Y have gradually undergone inactivating and loss mutations, from the effects of which the organism has largely been protected, through the continual presence of an X having normal ... allelomorphs. In other words, the Y has paid the penalty always exacted by the protection of continual heterozygosis, and the consequent absence of natural selection. The largely inert Y... must retain enough genes to allow it to act as the homologue of the X in segregation, if it is to persist at all, and, if any dominant genes exist or arise in it, which are advantageous to the sex in which Y occurs exclusively, they may be retained by natural selection' (Muller 1932 pp. 133-4). Graves (2006), with the aid of a great deal of study involving DNA polymorphisms, has taken Muller's prescient conclusions and confirmed or elaborated them: 'Thus many factors feed into equations describing the rate of degradation of the Y chromosome, and these make it difficult to predict how near to extinction the human Y really is. I challenge population and evolutionary geneticists to derive a meaningful model with predictive power. Essentially, the stochastic nature of many of the Y-major rearrangements and deletions on the negative side and acquisition of new male-advantage genes on the positive means that it is at the mercy of chance events. It seems unlikely that the human Y has achieved a stable state. It would take substitution of function of only a few genes to render the human Y completely redundant and permit its complete loss' (Graves 2006, p. 911). Engelstädter (2008 p. 957) is among those who have accepted Graves's challenge, and has concluded that 'mutations on the X chromosome can considerably slow down the [random loss of those chromosomes bearing the fewest deleterious alleles]. On the other hand, a lower mutation rate in females [XX] than in males [XY], background selection, and the emergence of dosage compensation are expected to accelerate the process.'

In flowering plants, where the basic form is hermaphrodite, selfing would be possible and likely if additional outbreeding mechanisms were not available. There are many of these; see Bateman (1952), Mayo (1983) and Leach and Mayo (2005) for discussion. I shall consider one system, gametophytically determined self-incompatibility (s.i.), to illustrate the special nature of the associated polymorphism.

In the single gene (locus) version of this system, there is one gene S , with alleles $S_1, S_2, S_3, S_4, \dots$ such that a female plant $S_i S_j$ can be pollinated and fertilisation effected by a pollen grain S_k where $i \neq j, i \neq k$ and $j \neq k$. Thus, all plants are heterozygous and the minimum number of alleles for a population to persist is 3. There is strong selection for equal frequencies of the three possible genotypes $S_1 S_2, S_1 S_3, S_2 S_3$ but in a population of finite size N there is a very low but non-zero probability ($< 3^{(1/2)^N}$) that only two genotypes will be represented in the offspring of any generation. Recognition of this fact led to the concept of the quasi-stable equilibrium (Ewens 1964) and a great deal of subsequent theory. If there are only 3 alleles and a new allele arises by mutation in a very large population, then it is at a substantial advantage and will rapidly increase in frequency to its equilibrium frequency of $1/4$. In most natural populations of plants possessing this outbreeding mechanism, very large numbers

of alleles are present. If populations are large, these alleles will not be lost, but with moderate population sizes, high mutation rates are required to maintain the numbers of alleles observed.

Multilocus versions of this system are known, representable as $S_{ij}S_{ik}$, $i = 1,2,3\dots$, $j \neq k = 1,2,3\dots$, whereby each allele of each gene behaves independently in pollen grain and style, so that pollination is possible unless all alleles in a pollen grain are represented in the style on which the pollen grain is lodged. Thus, the separate self-incompatibility genes in the multilocus system can behave as if they were not individually involved in a systematic disruption of panmixia; the only impossible genotypes are those homozygous for all self-incompatibility genes. Under such circumstances, fixation of one of the genes is highly likely unless all alleles have exactly the same selective value. Fixation at all but one of the loci brings the system back to that described above, which is highly resistant to fixation or disruption, except by other genes (or s-i alleles) which permit selfing, in which case fixation follows rapidly (Fisher 1941, Mayo and Leach 1987, Leach and Mayo 2005). Given these considerations, it is at first sight remarkable that many s.i. alleles are very old, in the sense that they appear to have appeared before the species that bear them (e. g. Richman 2000). Clark (1997, p. 7731) has noted that modelling of s.i. systems shows 'that the coalescence time of alleles [evolutionary time since divergence from some common ancestor] varies inversely with the rate of origination of novel functional alleles, and that for reasonable estimates of the rate of origination of new alleles, such extremely old polymorphisms are not unlikely'.

Sheltering of lethals is possible in the single locus system but much less likely in the multilocus systems (Leach *et al.* 1986). This is one small advantage the multilocus systems provide. It is of course entirely possible that quite different advantages accrue to multiple s.-i. loci, e.g. protein stability from duplication of active domains (Bhaskara and Srinivasan 2011); population-genetic modelling of such advantages awaits their demonstration at the cellular and organismal level.

It is difficult to explain how the more complex systems have evolved, just as with complex sex-determining systems. An initial advantage to a duplicated gene under strongly selective conditions (e.g. a population bottleneck) plus the regularity of Mendelian segregation in a subsequent expanded population might explain an individual case, but not the persistence of such systems in many very distantly related species.

5. Polymorphism at the DNA level

In the discussion so far, 'polymorphism' has simply referred to variants of a single Mendelian gene. These variants, however, may be anything from a change in just one DNA base in a sequence to a duplication of a whole structural gene or other lengthy sequence. Haemoglobin variants may, for example, arise from a single base change in one triplet codon. The S self-incompatibility locus alleles are much more complex variants. The human haptoglobin polymorphism has a number of alleles, of which one is a partial duplication of another (Smithies *et al.* 1962). Structural polymorphism of chromosomes has been shown to be very widespread (White 1973).

A restriction fragment length polymorphism (RFLP) is a polymorphism detected by DNA digestion with a so-called restriction enzyme. The polymorphism is a difference in the location of sites among two or more homologous DNA molecules (chromosomes) at which

the restriction enzyme cuts the DNA molecule. Thus, the alleles differ in length (size) and can be distinguished by gel electrophoresis.

Repeated sequences, called tandem repeats or satellite DNA, are classified by size: satellites are highly repeated, with the unit of repetition a thousand base pairs or more, so that the overall satellite is of the order of millions of base pairs in length; minisatellites are less repeated shorter sequences, 10-100 base pairs repeated sufficiently frequently give an overall length of thousands of base pairs; and microsatellites, which are shorter numbers of repetitions of sequences shorter than 10 base pairs. Satellites are found on Y chromosomes and near centromeres and telomeres, while microsatellites and minisatellites occur in the euchromatin of most eukaryotes. Minisatellites were the basis of most forensic DNA and much agricultural application until the recognition of how many SNPs there are.

Single nucleotide polymorphism, usually diallelic, and identified directly through DNA sequencing, has become very important since technologies for rapid DNA sequencing have become first feasible and now widely available. Development continues at a rapid rate (Rothberg *et al.* 2011). SNPs can occur anywhere in the genome, whether the DNA encodes structural genes, regulatory elements that are not translated, or DNA that is not transcribed. Although SNPs are now a vital tool for many kinds of investigation, their manifold effects are only beginning to be understood. For example, synonymous changes in codons (i.e. those that do not lead to a change in the encoded-for amino acids) in SNPs have been implicated in several diseases (Sauna and Kimchi-Scarfaty 2011).

Table 2 lists some of the different types of possible polymorphism. RNA different from transcribed DNA expectation (e.g. Li *et al.* 2011) has not been included because its phenotypic polymorphic outcomes are not clear.

DNA variation	Example of polymorphism
Whole chromosome	Sex determination
Whole chromosome inactivation	
Part of chromosome	Translocation, deletion, inversion, duplication
Whole gene	Translocation, deletion, inversion, duplication Protein polymorphism e. g. human haptoglobin
Whole gene inactivation	
Part of gene	Translocation, deletion, inversion, duplication
Part of sequence	
Restriction fragment length polymorphism (RFLP)	
Short tandem repeat (STR)	
Variable number tandem repeat (VNTR)/microsatellites & minisatellites	
Single nucleotide	

Table 2. DNA polymorphism

At every level of analysis from the chromosome down to the SNP, polymorphism can be used to investigate problems from the individual (e. g. risk prediction) to the population (e.g. the value and utility of ethnic group classification: Romualdi *et al.* 2002).

6. Polymorphism and quantitative variation

As already noted, the relationship between phenotype and polymorphism has been under investigation since the discovery of genetic polymorphism. Well characterised human polymorphisms such as the ABO blood groups have been shown to influence many quantitative traits, as well as being associated with disease as discussed in section 7 below. George and Elston (1987) provided a reliable method of analysis for use in small-scale human studies. Thus ABO influences serum cholesterol level (Mayo *et al.* 1969 and many subsequent studies) but the mechanism for this small effect is unknown. The same comment applies to the attraction for mosquitoes of some ABO phenotypes over others (Shirai *et al.* 2000).

Allen *et al.* (2010) have shown, through a meta-analysis and other studies of over 2,800,000 single nucleotide polymorphisms in over 180,000 individuals, that genetic variants in over 180 gene loci influence human height. There are no useful tools for modelling this kind of causation, other than the statistical methods built on the work of Fisher (1918), which cannot provide insight into physiological mechanisms.

In genetics applied to plant and animal improvement, useful methods can be developed which allow genetic improvement without a knowledge of physiological or molecular mechanism. Marker-assisted selection is the most important of these. The idea of using linkage disequilibrium for selection is not new. It was recognised early that detection of one or more markers associated with an increase in a desirable trait could allow selection for the trait using the markers. In this case, the markers would be associated with a chromosomal region influencing the trait of interest, such a region being called a quantitative trait locus (QTL). The method has, however, only become practicable with the discovery and mapping of thousands of DNA markers. Guimarães *et al.* (2007) give an account of the development of the field. For an example of industry application of marker-assisted selection, see Johnson and Graser (2010). They studied 12 commercial (GeneSTAR®) markers in populations from different breeds (temperate: Angus, Hereford, Murray Grey, Shorthorn; tropical: Santa Gertrudis, Belmont Red) and estimated effects associated with the markers 'that varied greatly across traits, suggesting large differences between the markers for their utility as selection tools in these populations' (p. 1917). If a QTL lies in a known gene in a known biochemical pathway, it can have a meaning other than its association with the trait of interest, but the utility of the association does not depend on such knowledge.

If markers have to be discovered, evaluated and applied for each trait, progress will be more certain than with traditional methods of animal breeding, but it will be costly and only slightly accelerated. Meuwissen *et al.* (2001) made a major advance in marker-assisted selection. Rather than simply search for individual markers of sufficient biological significance to be worth using in a breeding programme, they proposed a radically different approach using linkage disequilibrium between QTL and large numbers of markers across the genome without mapping the QTL. They introduced novel Bayesian statistical methods for estimating breeding value based on differing prior distributions of the effects of QTL and showed that high accuracy could be achieved. In 2001, cost-effective genotyping of very large numbers of markers was unavailable, but the tools used in the Human Genome project (section 7.2) radically reduced genotyping cost. SNP chips became available, so that Meuwissen *et al.*'s method, termed genomic selection, began to be implemented. Many dairy cattle breeders in advanced countries have adopted some of these new techniques. From

Meuwissen *et al.*'s work (and see also Goddard and Hayes 2007), genomic selection should almost double the rate of genetic improvement compared with traditional progeny testing. It has been claimed to be the most important advance in animal breeding since Henderson's (1953) Best Linear Unbiased Prediction. It will be applied to plant breeding.

7. Human examples

7.1 Association between polymorphism and disease

As cryptic human polymorphisms were identified, from the ABO blood groups in 1900 onwards, and at the same time many common diseases seen to 'run in families' were shown to have moderate to high heritability, it was natural that associations between phenotypes of a polymorphism and disease phenotypes would be sought. Discovery of such associations was expected to aid in the elucidation of the role of polymorphism in quantitative variation and might have been expected to allow risk prediction if very strong associations had been discovered. These analyses required the recognition that a disease state was a clinical definition related to a point on a scale of underlying liability (in the sense of Rendel 1967).

Detection of an association was relatively straightforward. Consider the 2 x 2 table:

		Polymorphic phenotype		Total
		A	not-A	
Disease state	Present	a	b	a+b
	Absent	c	d	c+d
	Total	a+c	b+d	n

Then the ratio $(a/c)/(b/d)$ gives the relative risk of disease in A and not-A persons. Table 3 gives a small number of well-established relative risks for blood group O as against blood group A. As noted by Bodmer and Bonilla (2008), these findings, though robust, have not been useful.

Disease	Relative risk
Duodenal ulcer	1.90
Gastric ulcer	1.19
Cancer of breast	0.92
Cancer of colon & rectum	0.90
Pernicious anaemia	0.80
Atherosclerosis	0.69

Table 3. Significant relative risk differences between ABO blood group phenotypes O and A for several diseases (extracted from Mayo 1978)

Once the methods of analysis had been developed, they were applied widely. The following data from two schizophrenia studies are reproduced from Mayo (1978):

Disease status	Schizophrenic		Non-schizophrenic control	
	O	A	O	A
Blood group phenotype				
South Australia	31	46	534	409
Lancashire	31	31	334	243

These data suggest a combined relative risk of 0.63 for schizophrenia between persons of blood group O and those of blood group A. This finding has not been confirmed in subsequent studies. In a similar way, many claimed associations of schizophrenia with other polymorphisms have been unsupported in subsequent studies (e.g. neuregulin 1 promoter polymorphism rs6994992: Crowley *et al.* 2008; D2 dopamine receptor gene *Taq1* polymorphism: Behravan *et al.* 2008; *Nogo* CAA 3' UTR insertion polymorphism: Gregório *et al.* 2005; Interleukin 10 gene promoter polymorphism: Jun *et al.* 2003; *NOTCH4* (CTG)_n polymorphism: Imai *et al.* 2001). Wray and Visscher (2010) give the best established associations as part of a thoughtful discussion of all of the issues in unravelling the genetical component of schizophrenia causation.

In some cases, for schizophrenia and for many other disorders, 'candidate genes' were considered likely to influence clinical outcomes. Other cases (like the ABO example above) reflected capability rather than expectation (see Edwards *et al.* 2011 for a broad discussion of this issue in biomedical research).

Even when a real and probably meaningful association is found, its interpretation can be complex (Table 4). Here, association of two behaviour-linked traits with a polymorphism means that both sets of behaviour would be needed to determine anything to do with mechanism, even if the finding were robust.

Caffeine intake (mg/d)	<i>ADORA2A</i> genotype		Odds ratio (95% CI)
	CC	CT TT	
(numbers of persons)			
All subjects (P for trend < 0.001)			
<100	150	100	1.0
100-200	261	129	0.7 (0.5, 1.0)
>200-400	1062	446	0.6 (0.5, 0.8)
>400	426	161	0.6 (0.4, 0.8)
Non-smokers (P for trend 0.03)			
<100	127	78	1.0
100-200	216	96	0.7 (0.5, 1.0)
>200-400	714	291	0.7 (0.5, 0.9)
>400	174	71	0.7 (0.4, 1.0)
Smokers (P for trend < 0.001)			
<100	23	22	1.0
100-200	45	33	0.8 (0.4, 1.7)
>200-400	348	155	0.5 (0.2, 0.9)
>400	252	90	0.4 (0.1, 0.7)

Table 4. Odds ratio of having the adenosine A2A receptor (*ADORA2A*) 1083TT genotype for caffeine intake among non-smokers and current smokers (extracted from Cornelis *et al.* 2007)

Failed replication of original associations has also occurred with many other important diseases. Diabetes is an example, and here it is noteworthy that newer techniques, in one particular case consideration of microRNAs, a class of regulatory molecule with broad but not yet fully defined or explicated function, can show how influence on disease is mediated by products of DNA sequences other than structural genes (Trajkovski *et al.* 2011). Basic

inflammatory mechanisms will be important in the underlying of disorders like diabetes that are closely related to inflammation (e.g. Liao *et al.* 2011), but this does not mean that associations will be found through 'candidate gene' polymorphisms.

New techniques for screening gene products rapidly increased the number of polymorphisms that could be investigated, making chance associations likely and increasing also the chance of finding real but meaningless associations brought about by linkage disequilibrium. These association studies are based on the idea that polymorphic alleles of a gene (or alleles of an unknown gene in strong LD with the test locus) can contribute to disease risk, i.e. there is a causal relationship. This contrasts with the rare variant hypothesis 'that a significant proportion of the inherited susceptibility to relatively common chronic diseases may be due to the summation of the effects of a series of low frequency dominantly and independently acting variants of a variety of different genes, each conferring a moderate but readily detectable increase in relative risk' (Bodmer and Bonilla, 2008, p. 696). As noted further by these authors (*loc. cit.*, p. 208) 'A critical feature shared by common and rare variants is that they do not give rise to a familial concentration of cases.' Yet many common chronic diseases show strong familial concentration, so these contrasted hypotheses cannot tell the whole story. (See also Mayo and Leach 2006.)

7.2 Human Genome Project

The success of the Human Genome Project was partly based on competition to develop new, faster, cheaper ways of sequencing DNA, both in the automation of the chemical analysis and in the statistical analysis that allowed sequences for large regions to be aligned after assembly from multiple overlapping shorter DNA sequences. Progress has continued to be rapid, so that sequencing costs have declined to the point where almost anyone in an OECD country could imagine having her or his own genome sequenced. This has brought new concerns, sharpening ethical concerns about the use of genetic information (for example, whether one should offer genetic risk diagnosis for a disease with no treatment, such as Huntington's) but also raises all the old ones: how should probabilistic risk estimates be used, e.g. those based on polymorphic disease associations? What level of risk requires medical or other intervention? And so on. These problems do not relate solely to polymorphism, of course.

More directly related to polymorphism is the strategy of large scale analysis. As the number of polymorphisms became almost indefinitely large through the Human Genome Project, and assay costs tiny relative to the cost of collecting the human disease and control samples, or other groups to be investigated, methods based on small samples collected on a 'one-off' basis and individual polymorphisms had to be discarded in favour of methods based on the whole genome, such as the genome-wide association study (GWAS). The GWAS has now been used for a large range of traits and diseases, from baldness (Hillmer *et al.* 2008; and see Abbasi 2011 for an insight into one of the genes that may be relevant) and eye colour (Liu *et al.* 2010) to neuroticism (Wray *et al.* 2008) and measured intelligence (e.g. Butcher *et al.* 2008).

Handsaker *et al.* (2011) give a good account of some of the strategic issues in population studies, and Allen *et al.* (2010), cited above, and the GWAS references in the previous paragraph set out what is necessary to conduct and then to combine many big studies using millions of SNPs. These strategic concerns do not, of course, mean that all the requirements

of a single study of a single polymorphism and a single disease, such as how inferences should be drawn from sample to population, must not still be met.

Table 5 lists some applications of polymorphism to genetic and other biological problems. It now seems clear that each individual carries some 2.5 million SNPs and substantial numbers of larger DNA polymorphisms. Post-HUGO international collaborations such as HapMap, set up to 'determine the common patterns of DNA sequence variation in the human genome and to make this information freely available in the public domain' (International HapMap Consortium 2003, p. 789), are beginning to assess this variability at higher levels of organisation.

Field	Application
Physiology	Gene and isozyme number
	Origin of cells and tissues
	Chromosome and gene inactivation
	Somatic cell hybridisation
Family studies	Linkage analysis
	Disease risk prediction
	Paternity testing
Population studies	Association of genes and diseases
	Genetic distance between groups
	Phylogeny
	Taxonomy
	Quantitative variation
	Pharmacogenetics/pharmacogenomics
	Sensory perception

Table 5. Application of polymorphism

8. Interaction

Interaction between a small number of factors, environmental or genetic, can readily be evaluated using standard statistical genetics. Table 4 is one example. Another is an association between a polymorphism of the fibrinogen β -chain gene and one influencing fibrinogen plasma concentration. Fibrinogen level is a risk factor for ischaemic heart disease (Woodward *et al.* 1998), yet the two polymorphisms, which are strongly associated, are not risk factors. Hence, the effect on heart disease must come about through mechanisms not yet elucidated (Vischetti *et al.* 2002).

Current knowledge suggests that interaction will be the norm for any carefully investigated trait, the understanding of the interaction depending on the level of investigation, which will almost always employ polymorphism as a tool. For example, Luo *et al.* (2001) used RFLP in rice to investigate heterosis in grain yield and its components and the contribution of epistatic interaction to heterosis. They obtained over 250 inbred lines from their F10 generation and mapped RFLP across them, assuming epistasis to be solely a digenic phenomenon. Growing the crossed varieties at 2 locations, they were able to assess the consistency of their findings. They found 30 quantitative trait loci (QTL) in toto: 7, 15 and 8 for panicles/plant, grains/panicle and 1000-grain weight, respectively. Just 1 QTL was the

same in both environments. Furthermore, only 8 of the 70 possible main effects associated were significant at the 0.1% level chosen by the authors. Taken together, these results imply very high levels of genotype \times environment interaction, and the authors concluded that '[o]verdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice'. Thus, RFLP had been used to derive a novel and important conclusion, which has been supported by other agricultural genetics studies, e.g. Barendse *et al.* (2007).

As noted above, the relationship between the determination of the heritable component of quantitative variation and polymorphism has not been fully elucidated. Insights will come from studies of regulatory DNA on a scale as yet barely envisaged (Frankel *et al.* 2011). Before then, however, new methods will be needed for dealing with multidimensional interaction on a scale never before attempted (Mayo, 2011).

9. Conclusions

The term 'polymorphism' is an old one which has survived many stages of reinterpretation and redefinition. Originally, it meant frequent or widespread but apparently meaningless variation in a population. Next, with the acceptance and development of the theory of Mendelian inheritance, it became a genetic concept, and with the coming of an understanding of stochastic forces in populations, genetic polymorphism was seen as something maintained by selection, or as transient while alleles were lost through chance variation. This was followed by the 'load' era, during which concerns were raised about the amount of deleterious mutation and the burden that this placed upon a population, even though the additive genetic variance in fitness, which is critical for the rate of change under natural selection (Fisher 1930), is proportional to the genetic load (Fraser and Mayo 1974). Indeed, in some formulations, load was seen as resulting from a population's deviation from an optimum genotype (Crow and Morton 1960). Now that we know that any individual human being may be polymorphic at more than a million SNP sites (loci) (e.g. Allen *et al.* 2010), with equivalent results for other outbreeding organisms, the idea of an optimum seems even more far-fetched.

The Human Genome project has revealed that humans have 'only' 20,000-30,000 structural genes, i.e. genes coding for proteins (International Human Genome Sequencing Consortium, 2004). However, much more of the DNA is translated, and its function is not yet understood, and it represents, at one level, many more 'genes', all interacting with the environment (e. g. Zhang *et al.* 2011). In addition, as noted above, the vast extent of DNA polymorphism, whereby even a SNP is a pair of segregating Mendelian alleles, means that polymorphism is the norm.

Looking back, it appears almost surprising that the effects of individual genes have been detectable and measurable so easily in so many cases. This has depended on the severity of deleterious mutations, on the visibility of many phenotypic variants, and on the keen eyes of medical practitioners and plant and animal domesticators and breeders. At a simple population genetic level, the analysis of Ewens and Thompson (1977) explains for fitness how an individual gene's effects are manifested. An important future task is to relate polymorphic genetic variation to phenotypic variation, whether for human diseases, production traits in livestock and crops, or fitness and other attributes of natural populations. Equally important, but not discussed in this chapter, is how polymorphism relates to gene regulation.

10. Acknowledgement

I thank CSIRO for my research fellowship and Carolyn Leach for improvements to the manuscript.

11. References

- Abbasi, A. A. 2011 Molecular evolution of HR, a gene that regulates the postnatal cycle of the hair follicle. *Scientific Reports* | 1 : 32 | DOI: 10.1038/srep00032. Accessed 13 July 2011.
- Allen, H. Lango and over 290 other authors. 2010 Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467 832-838.
- Barendse, W., Reverter, A., Bunch, R. J., Harrison, B. E., Barris, W. and Thomas, M. B. 2007 A validated whole-genome association study of efficient food conversion in cattle. *Genetics* 176 1893-1905.
- Behravan, J., Hemayatkar, M., Toufani, H. and Abdollahian, E. 2008 Linkage and association of DRD2 gene *Taq1* polymorphism with schizophrenia in an Iranian population. *Archives of Iranian Medicine* 11 252-256.
- Bhaskara, R. M. and Srinivasan, N. 2011 Stability of domain structure in multidomain proteins. *Scientific Reports* 1:40 | DOI:10.1038 | socp0040 1-8 accessed 22 July 2011.
- Bodmer, W. F. and Bonilla, C. 2008 Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics* 40 695-701.
- Bürger, R. 2000 *The Mathematical Theory of Selection, Recombination, and Mutation*. John Wiley & Sons, Chichester.
- Bürger R., and Hofbauer, J. 1994 Mutation load and mutation-selection-balance in quantitative genetic traits. *Journal of Mathematical Biology* 32 193-218.
- Burke, M. K., Dunham, J. P., Sharestani, P., Thornton, K. R., Rose, M. R. and Long, A. D. 2010 Genome-wide analysis of a long-term selection experiment with *Drosophila*. *Nature* 467 587-592.
- Butcher, L. M., Davis, O. S. P., Craig, I. W. and Plomin, R. (2008) Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes Brain Behavior* 2008 7 435-446.
- Clark, A. G. 1997 Neutral behavior of shared polymorphism. *Proceedings of the National Academy of Science* 94 7730-7734.
- Cornelis, M. C., El-Sohemy, A. and Campos, H. 2007 Genetic polymorphism of the adenosine A2A receptor is associated with habitual caffeine consumption. *American Journal of Clinical Nutrition* 86 240-244.
- Crow, J. F. and Morton, N. E. 1960 The genetic load due to mother-child incompatibility. *American Naturalist* 94 413-419.
- Crowley, J. J., Keefe, R. S., Perkins, D. O., Stroup, T. S., Lieberman, J. A. and Sullivan, P. F. 2008 The neuregulin 1 promoter polymorphism rs6994992 is not associated with chronic schizophrenia or neurocognition. *American Journal of Medical Genetics B Neuropsychiatric Genetics* 147B 1298-1300.
- Edwards, A. M., Isserlin, R., Bader, G. D., Frye, S. V., Willson, T. M. and Yu, F. H. 2011 Too many roads not taken. *Nature* 470 163-165.

- Engelstädter, J. 2008 Muller's ratchet and the degeneration of Y chromosomes: a simulation study. *Genetics* 180 957-967.
- Ewens, W. J. 1964 On the problem of self-sterility alleles. *Genetics* 50 1433-1438.
- Ewens, W. J. 2007 Fraser and the genetic load. Pp. 402-408 in *Fifty Years of Human Genetics a Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser*. (O. Mayo & C. R. Leach eds) Wakefield Press, Adelaide.
- Ewens, W. J. and Thomson, G. 1977 Properties of equilibria in multi-locus genetic systems. *Genetics* 87 807-819.
- Fisher, R. A. 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52 399-433.
- Fisher, R. A. 1922 On the dominance ratio. *Proceedings of the Royal Society of Edinburgh* 42 321-341.
- Fisher, R. A. 1930 *The Genetical Theory of Natural Selection*. Oxford University Press.
- Fisher, R. A. 1941 Average excess and average effect of a gene substitution. *Annals of Eugenics* 11 53-63.
- Ford, E.B. 1964 (4th edn 1975). *Ecological genetics*. Chapman and Hall, London.
- Frankel, N., Erezyilmaz, D. F., McGregor, A. P., Wang, S., Payre, F. & Stern, D. L. 2011 Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* 474, 598-603.
- Fraser, G.R. and Mayo, O. 1974. Genetical load in man. *Humangenetik* 23: 83-110.
- George, V. T. and Elston, R. C. 1987 Testing the association between polymorphic markers and quantitative traits in pedigrees. *Genetic Epidemiology* 4 193-201.
- Goddard, M. E. and Hayes, B. J. 2007. Genomic selection. *Journal of Animal Breeding and Genetics* 124 323-30.
- Graves, J.A.M. 2006. Sex chromosome dynamics and Y chromosome degeneration. *Cell* 12: 901-914.
- Gregório, S.P., Murya, F. B., Ojopia, E. B., Sallet, P. C., Morenoc, D. H., Yacubiana, J., Tavaresa H., Santos, F. R., Gattaza, W. F. and Dias-Netoa, E. 2005. Nogo CAA 3VUTR Insertion polymorphism is not associated with schizophrenia nor with bipolar disorder. *Schizophrenia Research* 75 5- 9.
- Guimarães, E. P., Ruane, J., Scherf, B. D., Sonnino, A. and Dargie, J. D. 2007. *Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish*. Food and Agriculture Organization of the United Nations, Rome. <http://www.fao.org/docrep/010/a1120e/a1120e00.htm> accessed 17 August 2011.
- JBS Haldane, J. B. S. 1940 The mean and variance of χ^2 , when used as a test of homogeneity, when expectations are small. *Biometrika* 31 346-355.
- Handsaker, R. E., Korn, J. M., Nemes, J. and McCarroll, S. A. 2011 Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nature Genetics* 43 269-278.
- Henderson, C. R. 1953. Estimation of variance and covariance components. *Biometrics* 9 226-252.
- Hillmer, A. M., Brockschmidt, F. F., Hanneken, S., Eigelshoven, S., Steffens, M., Flaquer, A., Herms, S., Becker, T., Kortüm, A.-K., Nyholt, D. R., Zhao, Z. Z., Montgomery, G. M., Martin, N. G., Mühleisen, T. W., Alblas, M. A., Moebus, S., Jöckel, K.-H.,,

- Bröcker-Preuss, M., Erbel, R., Reinartz, R., Betz, R. C., Cichon, S., Propping, P., Baur, M. P., Wienker, T. F., Kruse, R. & Nöthen, M. M. 2008. Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nature Genetics* 40 1279 - 1281.
- Imai, K., Harada, S., Kawanishi Y, Tachikawa H, Okubo T. and Suzuki T 2001 The (CTG)_n polymorphism in the NOTCH4 gene is not associated with schizophrenia in Japanese individuals. <http://www.biomedcentral.com/content/backmatter/1471-244X-1-1-b1.pdf> . Accessed 21 June 2011.
- International HapMap Consortium 2003. The International HapMap Project. *Nature* 426, 789-796.
- International Human Genome Sequencing Consortium. 2004 Finishing the euchromatic sequence of the human genome. *Nature* 431 931-945.
- Johnston, D. J. and Graser, H.-U. 2010. Estimated gene frequencies of GeneSTAR markers and their size of effects on meat tenderness, marbling, and feed efficiency in temperate and tropical beef cattle breeds across a range of production systems. *Journal of Animal Science* 88 1917-1935
- Jun, TY, Pae CU, Kim KS, Han H and Serretti A. 2003 Interleukin-10 gene promoter polymorphism is not associated with schizophrenia in the Korean population. *Psychiatry and Clinical Neuroscience* 57 153-159.
- Leach, C. R. & Mayo, O. (2005) *Outbreeding mechanisms in flowering plants: an evolutionary perspective from Darwin onwards*. Stuttgart, J. Cramer (E. Schweizerbart'sche).
- Leach, C.R., Mayo, O. and Morris, M.M., (1987). Linkage disequilibrium and gametophytic self-incompatibility. *Theoretical and Applied Genetics* 73: 102-112.
- Li, M., Wang, I. X., Li, Y., Bruzel, A., Richards, A. L., Toung, J. M. and Cheung, V. G. 2011 Widespread RNA and DNA sequence differences in the human transcriptome. *Scienceexpress* 10.1126science.1207018. Accessed 25 May 2011.
- Liao X. , Sharma, N., Kapadia, F., Zhou, G., Lu, Y., Hong, H., Paruchuri, K., Mahabeleshwar, G. H., Dalmas ,E., Venteclef, N., Flask, C. A., Kim, J., Doreian, B. W., Lu, K. Q., Kaestner, K. H., Hamik, A., Clément, K. and Jain, M. K. Krüppel-like factor 4 regulates macrophage polarization. *Journal of Clinical Investigation*, 2011; DOI: 10.1172/JCI45444
- Liu, F., Wollstein, A., Hysi, P. G., Ankra-Badu G. A., Spector, T. D., Park, D., Zhu, G., Larsson, M., Duffy, D. L., Montgomery, G. W., Mackey, D. A., Walsh, S., Lao, O., Hofman, A., Rivadeneira, F., Vingerling, J. R., Uitterlinden, A. G., Martin, N. G., Hammond, C. J., Kayser, M. 2010: Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genetics* 6(e1000934):1-15, accessed 20 July 2011.
- Luo, L. J., Lia, Z.-K., Mei, H. W., Shu, Q. Y., Tabien, R., Zhong, D. B., Ying, C. S., Stansel, J. W., Khush, G. S., and Paterson, A. H. 2001 Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics* 158 1755-1771.
- Maynard Smith, J. and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. *Genetical Research* 23 23-35.

- Mayo, O. 1971. Rates of change in gene frequency in tetrasomic organisms. *Genetica* 42 329-337.
- Mayo, O. 1976. Neutral alleles at X-linked loci: a cautionary note. *Human Heredity* 26: 263-266.
- Mayo, O. 1978. Polymorphism, selection and evolution, in *The Biochemical Genetics of Man* (2nd edition).
- Mayo, O. 1983. *Natural Selection and Its Constraints*. Academic Press, London.
- Mayo O. 2007. The rise and fall of the common disease-common variant (CD-CV) hypothesis: how the sickle cell disease paradigm led us all astray (or did it?) *Twin Research and Human Genetics* 10 793-804.
- Mayo, O. 2011 Interaction between genotype and environment: a tale of two concepts. *Transactions of the Royal Society of South Australia* 135 113-123.
- Mayo, O., Fraser, G. R. and Stamatoyannopoulos, G. 1969. Genetic influences on serum cholesterol in two Greek villages. *Human Heredity* 19 86-99.
- Mayo, O. and Leach, C.R. (1987). Stability of self-incompatibility systems. *Theoretical and Applied Genetics* 74: 789-792.
- Mayo O & CR Leach (2006) Are common, harmful, heritable mental disorders common relative to other such non-mental disorders, and does their frequency require a special explanation? *Behavioral and Brain Sciences* 29 415-416.
- Meuwissen, T. H. E., Hayes, B. J. and M. E. Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 1819-1829.
- Morton, N. E. 2007 Genetic loads half a century on. Pp. 431-435 in *Fifty Years of Human Genetics a Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser*. (O. Mayo & C. R. Leach Eds) Wakefield Press, Adelaide.
- Muller, H. J., 1914 A gene for the fourth chromosome of Drosophila. *Journal of Experimental Zoology* 17 326-328.
- Muller, H. J. 1932 Some genetic aspects of sex. *American Naturalist* 66 118-138.
- Muller, H. J. 1950 Our load of mutations. *American Journal of Human Genetics* 2 111-176.
- Nagylaki, T. 1992 *Introduction to Theoretical Population Genetics*. Springer-Verlag, Berlin.
- Rendel, J. M. 1967 *Canalisation and Gene Control*. London, Logos Press.
- Richman, A. D. 2000 S-allele diversity of *Lycium andersonii*: implications for the evolution of S-allele age in the Solanaceae. *Annals of Botany*, 85 (Supplement A) 241-245.
- Romualdi, C., Balding, D., Nasidze, I. S., Risch, G., Robichaux, M., Sherry, S. T., Stoneking, M., Batzer, M. A. and Barbujani, G. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Research* 12: 602-612.
- Rose, C. J., Chapman, J. R., Marshall, S. D. G., Lee, S. F., Batterham, P., Ross, H. A. and Newcomb, R. D. 2011 Selective sweeps at the organophosphorus insecticide resistance locus, *Rop-1*, have affected variation across and beyond the α -esterase gene cluster in the Australian sheep blowfly, *Lucilia cuprina*. *Molecular Biology and Evolution* 28 1835-1846.
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A.,

- Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fianza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T. & Bustillo, J. 2011 An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475 348-352.
- Sauna, Z. E. and Kimchi-Sarfaty, C. 2011 Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics* 12 683-691.
- Shirai, Y., Kamimura, K., Seki, T. and Morohashi, M. 2000. Proboscis amputation facilitates the study of mosquito (Diptera: Culicidae) attractants, repellents, and host preference. *Journal of Medical Entomology* 37 637-639.
- Smithies, O., Connell, G. E. and Dixon, G. H. 1962 Chromosomal rearrangements and the evolution of haptoglobin genes. *Nature* 196 232-236.
- Sved, J. A. 1968. The stability of linked systems of loci with a small population size. *Genetics* 59 543-563.
- Sved, J. A. 2007 Deleterious mutations and the genetic load. Pp. 461-467 in *Fifty Years of Human Genetics a Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser*. (O. Mayo & C. R. Leach Eds) Wakefield Press, Adelaide.
- Tills, D., van den Branden, J. L., Clements, V. R. and Mourant, A. E. 1971 The world distribution of electrophoretic variants of the red cell enzyme adenylate kinase. *Human Heredity* 21 302-331.
- Trajkovski, M., Hausser, J., Soutschek, J., Bhat, B., Akin, A., Zavolan, M., Heim, M. H., & Stoffel, M. 2011 MicroRNAs 103 and 107 regulate insulin sensitivity. *Nature* 474 649-654.
- Vischetti, M., Zito, F., Donati, M. B., and Iacoviello, L. 2002 Analysis of gene-environment interaction in coronary heart disease: fibrinogen polymorphisms as an example. *Italian Heart Journal* 3 18-23.
- Watson, H. W. and Galton, F. 1875. On the probability of the extinction of families. *Journal of the Anthropological Institute of Great Britain* 4, 138-144.
- White, M. J. D. 1973 *The Chromosomes*. London, Chapman and Hall.
- Woodward, M., Lowe, G. D. O., Rumley, A. and Tunstall-Pedoe, H. 1998 Fibrinogen as a risk factor for coronary heart disease and mortality in middle-aged men and women The Scottish Heart Health Study. *European Heart Journal* 19 55-62.
- Wray, N. R., Middeldorp, C. M., Birley, A. J., Gordon, S. D., Sullivan, P. F., Visscher, P. M., Nyholt, D. R., Willemsen, G., de Geus, E. J. C., Slagboom, P. E., Montgomery, G. W., Martin N. G. and Boomsma, D. I. 2008 Genome-wide linkage analysis of multiple measures of neuroticism of 2 large cohorts from Australia and the Netherlands. *Archives of General Psychiatry* 65 649-658.
- Wray, N. R. and Visscher, P. M. 2010. Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophrenia Bulletin* 36: 14-23.
- Wright, S. G. 1930 Evolution in Mendelian populations. *Genetics* 16 97-159.
- Zaykin, D. V., Pudovkin A. and Weir B. S. 2008. Correlation-based inference for linkage disequilibrium with multiple alleles. *Genetics* 180 533-545.

- Zhang G., Chen, X., Chan L., Zhang M., Zhu, B., Wang, L., Zhu, X., Zhang, J., Zhou, B. and Wang, J. 2011 An SNP selection strategy identified IL-22 associating with susceptibility to tuberculosis in Chinese. *Scientific reports* |1:20| DOI: 10.1038/srep00020 accessed 29 June 2011.
- Zhang, L., Hou¹, D., Chen, X., Li, D., Zhu, L., Zhang, Y., Li, J., Bian, Z., Liang, X., Cai, X., Yin, Y., Wang, C., Zhang, T., Zhu, D., Zhang, D., Xu, J., Chen, Q., Ba, Y., Liu, J., Wang, Q., Chen, J., Wang, J., Wang, M., Zhang, Q., Zhang, J., Zen, K. and Zhang, C-Y. 2011 Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Research* advance online publication 20 September 2011; doi:10.1038/cr.2011.158 accessed 26 September 2011.

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.