THE CONCEPT OF THE GENE

E. M. EAST

Bussey Institution, Harvard University, Boston, Massachusetts

East, E. M. 1929 The concept of the gene. *Proceedings of the International Congress of Plant Sciences, Ithaca New York, August 16–23, 1926,* vol. 1. Menasha, WI: George Banta Publishing Co. pp. 889–895.

Electronic Scholarly Publishing Project

Foundations Series: Classical Genetics

Series Editor: Robert J. Robbins

The ESP Foundations of Classical Genetics project has received support from the ELSI component of the United States Department of Energy Human Genome Project. ESP also welcomes help from volunteers and collaborators, who recommend works for publication, provide access to original materials, and assist with technical and production work. If you are interested in volunteering, or are otherwise interested in the project, contact the series editor: rrobbins@fhcrc.org.

Bibliographical Note

This ESP publication is a newly typeset, unabridged version, based on the original publication. All footnotes and endnotes are as they appeared in the original work.

Production Credits

Scanning of originals:	ESP staff
OCRing of originals:	ESP staff
Typesetting:	ESP staff
Proofreading/Copyediting:	ESP staff
Graphics work:	ESP staff
Copyfitting/Final production:	ESP staff

East, E. M. 1929 The concept of the gene. *Proceedings of the International Congress of Plant Sciences, Ithaca New York, August 16–23, 1926,* vol. 1. Menasha, WI: George Banta Publishing Co. pp. 889–895.

THE CONCEPT OF THE GENE[1]

E. M. EAST

Bussey Institution, Harvard University, Boston, Massachusetts

NEARLY FIFTEEN YEARS AGO[2] I attempted to defend the thesis that the Mendelian method of recording the facts of inheritance was simply a notation useful as a description of physiological facts. The argument was an elaboration of the proposition that the germ-cell unit of heredity, the gene, was an abstract, formless, characterless concept used for convenience in describing the results of breeding experiments. It was the ghost of an entity which might later be clothed with flesh, but its usefulness at the time was due to its adaptability to mathematical treatment. By postulating that the results derived from controlled matings were due to the activities of definite germ-cell units which could be manipulated arithmetically, investigators were able to formulate new experimental tests, and thus to open the way to further discovery; but these units could be given no intelligible interpretation in terms of geometry, chemistry, or physiology.

There is no reason today why the statements made at that time should be repudiated; rather should they be re-emphasized and made more rigorous, for, although numerous plants and animals have told us

1 Presented before the International Congress of Plant Sciences, Section of Genetics, Ithaca, New York, Aug. 19, 1926.

2 East, E. M., The Mendelian notation as a description of physiological facts. *Amer. Nat.* 46: 633–655. 1912.

something of their heritage during the interim, and part of the phantasmagoria of the gene is now history, there is still a tendency to visualize it more concretely than is right and proper. And there is this further justification for reflecting on the groundwork of genetical theory from time to time; the difficulty which so many biologists experience in viewing biological phenomena with mathematical spectacles largely accounts for the limited diffusion of genetical philosophy outside of the ranks of specialists.

It is customary to regard biology as a science that moves slowly as compared with one of the supposedly more exact sciences such as chemistry. Probably this estimate is correct; yet the basic chemical concept, the modern atom, which is quite analogous to the genetic unit, remained a characterless mathematical notation for almost a century in spite of the alluring images drawn by Lord Kelvin and others. Although there had been atomic concepts in classical times, the modern atom was called into use by the discovery of the Law of Definite and Multiple Proportions, and served chemistry in a purely mathematical way until the discovery of radium. Since then, in a period practically coextensive with the history of genetics, it has been given such a quota of qualities by Bohr and his co-workers that it can now take its place as a recognizably respectable citizen. During this same time-phase, the genetic unit unquestionably has reached nearly as high a social status.

A geneticist may make such a statement as this, confident that it is just; yet if he is not carried away by enthusiasm, he must also admit that this central concept, the theory of the gene, has not been incorporated into general biology as part of its legal estate. One should expect the period of probation usually required by science for new ideas, but in the case of the gene concept the time has been overly long. Yet it is not particularly strange that this should be so. From the time when Adam began to name the animals, as duly recorded in the second chapter of Genesis down to the present, the greater part of all biological work has been historical and descriptive. Occasionally provisional causes for known facts have been adopted as a basis for further investigation, hut the premises invariably have been simple and the conclusions direct. The evolution theory is a case in point. The supporting evidence is voluminous and its character varied, but the theory itself is free from intricacy or subtlety. The ideas involved in the term gene, on the other hand, are both complex and abstract. Because of the novelty of such conceptual notions, therefore, the biologist is inclined to approach them cautiously, sometimes even electing an attitude of arrogant distrust as a defense reaction.

The student of the living is not wholly to be blamed for his reactions. His position is not an easy one. The worker in the physical

sciences, who has been making use of abstract concepts from the time of Hipparchus, has been somewhat inclined to patronize the biologist because the latter has occupied his mind with percepts or imagery familiar to the senses, to which the naive investigator can cling with perfect faith in its reality. But the physicist is not nearly so tough-minded an individual as he professes to be. He is merely fortunate in being able to segregate his activities. He can do sound scientific work requiring the highest type of objective reasoning, and at the close at his working day lock up his professional personality in his laboratory, don a different ego with his dinner coat, and wander forth to dabble with theological dogma or to search for ectoplasm. The biologist has no such advantages. He is bound up with the problems of life at all times. He is psychiatrist to this lunatic asylum of the universe, and is familiar with all the crude absurdities with which man likes to deceive himself, including even the reason why the physicist likes to titillate his emotions in his idle hours. Thus it is difficult for him to find an emotional outlet of his own unless he abandons the proved tools with which he has successfully delved into the unknown and goes in wholly for entelechies or similar intentionally untestable figments of the imagination. The result is that his emotions and his intellect are forever embroiled. He hopes against hope that his old heritage of beliefs is true, that environment is all-powerful, that free-will prevails, that man is created in the image of God and is only a little lower than the angels. And he keeps on, a pathetic figure, proving that all the old folkways are myths, in spite of his desires and hopes. He does all this but the conduct wearies him and makes him slow, slow to accept his own facts, slow to push them to their logical end. I can see no other reason for the dozens of recent biologies, particularly genetic biologies, which have started bravely to build a scientific edifice to house man's beliefs, but have finished by decorating it with so many saintly old *mores* that it looked like a cathedral after all. I can see no other reason for the hesitation about accepting the theory of the gene. It is first the novelty of the thing, and second the effect it has on the old folkways.

In addition to the general difficulty experienced by biologists in accepting the new when derived from a more or less materialistic hypothesis, there is the specific embarrassment, which all of us experience to a greater or less degree, of distinguishing between the fiction postulated as an aid to gaining established fact and the facts themselves. We come to believe in our fictions and hesitate to give them up when they are no longer necessary. Science is almost as prone as theology to glorify its fictions into dogmas, though usually they do not stand in the way of progress for such lengths of time. Perhaps, therefore, it would be a service to genetics if the story of its building

were called to mind every little while, that we may actually see what has been useful as scaffolding and what remains as structure.

Mendel's postulate of a germ-cell unit, which we may call the gene in order not to change the terminology, was not a novelty. Quantitative science requires units. They have been used from time immemorial. As soon as biologists began to speculate on the possibility of reducing the phenomena of heredity to law, therefore, units of description were proposed. At about the same time appeared the suggestions of Nägeli, Spencer, Darwin, Weismann and Mendel. *Mendel made his units useful.*

It appeared that the pea experiments could be interpreted by units from a duplex organization consisting of homologous pairs in the somatic cells, which could undergo varied associations with other units without being modified and afterwards appear in the germ cells in simplex organization produced by the permutations and combinations of one member of each homologous pair. Mendel's generalizations were three in number, the segregation of homologous factors without change of identity in the formation of the germ cells, the recombination of the products of this segregation in all possible combinations, and the formation of zygotes by random matings among the gametes thus produced.

These three laws have now been shown to be special cases of more general phenomena. The law of segregation retains the original idea of concrete units undisturbed in their own identity by association with other units in the germplasm; but the idea of a double set of hereditary factors, serially homologous, becoming two separate sets by the required choice of *one* of each pair of factor mates, no longer tells the whole story. The chromatin has been proved to be the gene carrier; its distribution gives us the distribution of the genes. And, like the activities of human beings, the behavior of chromatin is not always according to the regular or accepted mode.

Nor is the second law of Mendel, chance recombination of genes, a complete expression of the facts. Generally speaking, the individual chromosomes behave as if they were more or less independent transportation systems for packets of genes, but these packets may be broken up and interchanged according to regular rules, the rules themselves being subject to the influence both of external conditions and of hereditary factors. It is even possible that several chromosomes, ordinarily segregating independently, may tend to keep their maternal and paternal associations. At least one may so interpret some recent

work of Gates[3] on mice where a series of dominant characters in the house mouse which ordinarily assort independently showed a tendency to linkage when opposed to recessive characters in the Japanese waltzer.

Finally, it would not be at all strange if the random mating of gametes to form zygotes is also a special case. Clearly random mating is what usually occurs; but so many cases are now known where there is a differential *opportunity* for gametic unions that it is quite probable that it will soon be impossible to draw a sharp line between equal and unequal opportunities for fertilization.

These cases where the original Mendelian laws have broken down have been of paramount importance to genetics. Progress is nearly always due to the analysis of exceptional phenomena. A complete list would practically be the history of the science.

The interaction of two independently inherited factors to bring about a single visible effect resulted in an early extension of theory. A series of dominants, each one epistatic to the next lower, called forth the Presence and Absence Theory. Probably no geneticist ever believed in quite such a crude Presence and Absence Theory as Morgan has made out; but that is of little consequence. The theory served a purpose and gave way to a clearer one after analysis of exceptional cases had shown that an individual heterozygous for a factor is different from the individual that is haploid for it, and that in general both dominants and recessives function actively, though either may be inactivated by the presence of other genes and possibly by changes in external conditions.

Again, these "irregular" cases have told us much about the relation of genes to each other. No one now attributes a single specific effect to one gene or interprets a particular organic character as the result of one gene's activity. Single genes affect many characters; each character is built up by the action of many genes. These facts have been proved most beautifully by Bridges' work on sex in *Drosophila melanogaster*.

Perhaps the greatest, or at least the most spectacular, result of modern pedigree culture analysis supplemented by cytological investigations however, is the contribution to our knowledge of the architecture of the germ-cells. Proof of the linear arrangement of the genes in the chromosomes, which was clinched by non-disjunction and reduplication phenomena, was an astounding achievement of inductive experiment. The basis of all these discoveries is "crossover" frequency, or breaks in linkage. Let us discuss this genetic tool. It is perhaps more

3 Gates, W. H. The Japanese waltzing mouse: its origin, heredity, and relation to the genetic characters of other varieties. *Pub. Carn. Inst. Wash.* No. 337, 1926.

worth while than the discussion of any other genetic tools or attainments, since a little meditation brings out clearly that we have after all reached only a relative truth. The gene even now has no very concrete meaning. It is a slowly shifting, imaginative figure, clothed indeed with some known qualities, but still vague, very vague.

Take first the question as to whether a given effect is due to two genes or to one gene. A plant, let us say, has a red flower and a red fruit. Varieties exist with white flowers and white fruits. Ordinarily, by crossing, one can find out shortly whether the effect comes from one gene or from two, and if from two genes whether they lie in separate chromosomes. But the matter is largely a question of frequency. If the cross is *AB* × *ab* and there is a five per cent crossover, sufficient *Ab* and *aB* gametes can be traced to satisfy our doubts. But suppose the crossover is one one hundredth of one per cent. Is it a crossover? or is it a mutation? If both *Ab* and *aB* gametes can he traced we call it a crossover: if only *Ab* or *aB* gametes can be traced, which is quite likely to be the case—especially when investigating megasporogenesis—we should call it a mutation.

Multiple allelomorphism is another illustration of the artificiality of the system. When a change occurs, which when crossed back with the original stock gives only the two classes, we say a mutation has taken place. A third, a fourth, any number of changes occur in this character. When any two are crossed and only these two recovered in the F_2 generation, we say that a series of multiple allelomorphs has been established. Such usage is logical and desirable but it by no means proves that chemical rearrangement has occurred in each case at one definite place in the substance of the chromosome. Morgan[4] believes that it does, but his point is not well taken. He says:

> It might be claimed that the phenomenon of multiple allelomorphism is an expression, not of changes in the same locus, but changes in neighboring loci that are so close together that crossing-over never occurs between them. Suppose, by way of example, a mutation took place near the white locus that gave the eye color eosin. If the mutation occurred in a chromosome that had already a white gene, then the new eye color would be due to the combined action of white and eosin. If eosin arose in this way, then, when such an eosin fly is crossed to the original white stock it should give white, since the effect of the new recessive eosin in one chromosome of the F_1 is cancelled by the effect of the normal allelomorph of eosin. In fact, the combination gives an eye color that is not white but is intermediate between white and eosin. If, on the other hand, the mutation to eosin

4 Morgan, T. H., Bridges, C. B., and Sturtevant. A.H., The genetics of Drosophila. *Bibliographica Genetica II*, 1925. 262 p. (See p. 37.)

> occurred in an X-chromosome not having a white gene, but its normal wild-type allelomorph, then whenever an eosin fly is crossed to a white fly the female offspring should have red eyes, since each X carries the normal allelomorph of the recessive gene (white or eosin) of the other X. But in fact, an eye-color not red but intermediate between white and eosin results, demonstrating again that eosin and white are not closely linked recessive mutants. The same arguments apply to each of the other nine members of the same series. In the light of these facts it is surprising that statements are still made that the evidence of multiple mutation in the same locus is not established.

I do not believe that this argument is cogent. There have been twelve mutations at the so-called "white locus" in *Drosophila melanogaster*. Let us assume that a change has taken place in each case at a different "spot" in a linear series of twelve "spots" all so close together, or for some reason bound so strongly together, that crossing-over never takes place. A change takes place at No. 1 link which produces white, a change takes place at No. 2 link which produces pink, a change takes place at No. 3 link which produces eosin. Now Morgan says that "if eosin arose in this way, then, when such an eosin fly is crossed to the original white stock it should give white, since the effect of the new recessive eosin in one chromosome of the F_1 is cancelled by the effect of the normal allelomorph of eosin." This is either imagining too much or not enough. The change in the No. 1 link gave white without respect to what remained. The change in No. 3 link gave eosin without respect to what remained. A cross between the two would give whatever the combined effects of the haploid white and the haploid eosin happened to be, and only these two forms would be recovered in the F_1 generation, since no crossing-over has been postulated.

Such a question holds little interest for theoretical genetics at present, but it does bear on the nature of the gene. Morgan has endeavored to make a determination of gene size at a given period in gametogenesis. He has done this by calculating the volume of the chromosomes at this phase and then taking as the unit size that portion of the chromosome giving the smallest crossover value. There is difficulty in making both calculations, particularly the second; but by taking approximately the lowest value of the crossover curve, he arrives at a value which is one-fifth of a unit distance. The diameter of the gene, thus calculated, turns out to be 60/1000 of a micron. The number of genes he estimates at 2000.

The gene, when estimated in this manner, proves to be but little larger than the calculated size (also an interpretation, of course) of the molecule of haemoglobin. It is an ingenious and valuable computation;

but if we are not careful, it may lead our ideas astray. The estimate of the total number of genes is based on the assumption that all genes are alike, that our unit is a spatial unit instead of a genetic unit. The *Drosophila* chromosome "map" contains vacant spaces. Genes are more frequent in some parts than in others. It is assumed from this fact that mutations have occurred—or have been discovered—more frequently in some parts of the chromosome than in others. But it is also reasonable to suppose that genes vary in size. Separate genes may not have been discovered because non-crossover space is large in some regions and small in other regions.

Now such speculation is useless unless it serves to aid in holding our minds open for future eventualities. At present too rigid a visualization of the gene is not wise because of its effect on the reception of data concerning the frequency of mutation. It is very fortunate that the gene as we have thus far clothed it with flesh has a sufficiently high degree of stability to be extensively useful. Naturally the ultimate unit of any nomenclature must be stable, but sometimes it is necessary to utilize several units to serve all purposes. Synthetic and analytic chemistry finds the molecule useful as a unit, though molecules differ greatly in size and in stability. For other purposes the atom is valuable. In final analysis, however, recourse has been made to a still less variable unit, the electron. Is genetics going to find it necessary to deal with a lesser unit of heredity in order to gain stability? This question is difficult to answer. Eyster's work deals with an extremely variable gene. It may be necessary to postulate a new unit to deal with it or with similar phenomena effectively. But perhaps, with the analogy of chemistry before us, we can make the old conception serve, if we keep in mind that we know as yet neither the upper nor the lower limit for gene size, neither the upper nor the looser limit for gene stability and that we have drawn our pictures merely by the aid of crossover values. It is reasonable to suppose that genes are subject to the laws of chemistry. Let us assume that they are organized like molecules. It would follow that some gene-molecules are enormous as compared with others, that some are highly stable and others relatively unstable; and it might even follow that there is a correlation between size and instability. Furthermore, since chemical molecules can undergo various reorganizations without loss of substance, perhaps gene-molecules can do the same. These reorganizations may or may not be equally possible in both directions: that is to say, each state may not have the same coefficient of stability. Again, on chemical analogy, only a limited number of reorganizations should fit the communistic purposes of the organism; the majority of mutations should be lethal. And finally, since increasing molecular complexity decreases the

available directions of change, though it increases the number of possible permutations, orthogenesis is understandable. This last sentence is merely a statement of the fact that although highly organized molecules, of the aromatic series or of the aliphatic series, let us say, have many possibilities of easy reorganization, these reactions are more limited in direction because of the large and stable molecular nucleus than are of those of certain simpler compounds.

We arrive, therefore, at the same port from which we departed when our discussion began. The genes are units useful in concise descriptions of the phenomena of heredity. Their place of residence is the chromosomes. Their behavior brings about the observed facts of genetics. For the rest, what we know about them is merely an interpretation of crossover frequency. In terms of geometry, chemistry, physics or mechanics, we can give them no description whatever.