



Of genes and genomes: challenges for the twenty-first century

Richard A. Jorgensen*

Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y Estudios Avanzados, Irapuato, Guanajuato, México

*Correspondence: rajorgensen@langebio.cinvestav.mx

WHAT IS A GENE?

The past hundred years of genetics research produced astonishing advances in knowledge of genes and genomes, and yet full understanding of the nature of the gene still remains a major challenge. To explore why this is so, and to frame the question in twenty-first century terms, it is interesting to consider the mid-twentieth century thoughts of the maize geneticist L.J. Stadler.

In a seminal 1954 article, entitled “The Gene,” Stadler explored the nature of the gene by applying to it the “operational viewpoint,” an approach borrowed from modern physics that is based on the principle that: “an object or phenomenon under experimental investigation cannot usefully be defined in terms of assumed properties beyond experimental determination, but rather must be defined in terms of the actual operations that may be applied in dealing with it” (Stadler, 1954). Thus, he asked: “What is a gene in operational terms? In other words, how can we define the gene in such a way as to separate established fact from inference and interpretation?”

Stadler’s answer was that “*operationally, the gene can be defined only as the smallest segment of the gene-string that can be shown to be consistently associated with the occurrence of a specific genetic effect* [emphasis added].” Equally important to Stadler were the ways by which the gene cannot be defined: (1) “it cannot be defined as a single molecule, because we have no experimental operations that can be applied in actual cases to determine whether or not a given gene is a single molecule”; (2) “it cannot be defined as an indivisible unit, because, although our definition provides that we will recognize as separate genes any determiners actually separated by crossing over or translocation, there is no experimental operation that can prove that further separation is impossible”; and (3) “for similar reasons, it cannot be defined as the unit of reproduction or the unit of action of the gene-string, nor can it be shown to be delimited from neighboring genes by definite boundaries.”

Stadler’s definition and exclusions still apply today: genes are clearly not single molecules; chromosomes are. Chromosomes are collections of many genes, and delimiting genes from neighboring genes on a chromosome remains a tremendous, unsolved challenge, even with the complete sequence of a chromosome in hand. Moreover, epigenetics has shown that phenotypically defined “mutations” can have a physical basis other than the alteration of DNA sequence; rather, they may sometimes turn out to be “epimutations” based on modification states of DNA and chromatin proteins.

Thus, just as in Stadler’s day, “questions concerning the undetermined properties” of the gene remain “the all-important questions that we hope ultimately to answer by the interpretation of the experimental evidence and by the development of new experimental operations.”

Importantly, Stadler distinguished between the operationally defined gene and the *hypothetical gene*, pointing out that use of the term “gene” in the literature sometimes referred to the operational gene, sometimes the hypothetical gene, and sometimes “a curious conglomeration of the two,” an observation that applies to today’s literature as much as it did in 1954. The difference between the two concepts is straightforward: “The operational definition merely represents the properties of the actual gene, so far as they may be established from experimental evidence by present methods. The inferences from this evidence provide a tentative model of the hypothetical gene, a model that will be somewhat different in the minds of different students of the problem and will be further modified in the light of further investigation.”

THE GENE TODAY

Uses of the term “gene” today are many and varied, and unfortunately, often careless and incorrect. For instance, it is not uncommon for the term “gene” to be used to refer only to its protein coding sequences, thereby unconsciously redefining the gene

merely in terms of its ultimate output, and failing to understand that the gene (in modern, molecular terms) is a unit comprised of many interdependent elements, including all those elements *in cis* that are necessary for the normal operation of a given gene. This broader, more inclusive definition makes the “delimiting” of the gene in molecular terms very difficult, of course, because it requires not only we be able to identify accurately all intron–exon boundaries, but also all transcriptional control elements that determine when and where a gene is expressed, as well as non-protein coding signals in the DNA and the expressed RNA, such as transcription initiation sites, polyadenylation signals, alternative splicing signals, and translational control signals.

And of course today we are more aware than ever that many genes encode only RNA molecules as their functional products, as illustrated by the tremendous diversity of microRNAs that are found in the genome (often referred to as “non-coding” RNAs, though quite demonstrably they encode information that regulates the expression of other genes through RNA turnover and translational control).

Clearly then, understanding the complete coding capacity of a genome is a leading grand challenge for genetics and genomics in the early twenty-first century, and it seems likely to remain so for some time to come.

CHROMOSOMES

No less challenging than understanding the nature of the gene and the complete coding capacity of genomes is understanding the mechanisms that determine the integrity and dynamic behaviors of chromosomes and genomes, including functional structures like telomeres and centromeres, and dynamic processes such as replication, recombination, repair, condensation, localization, mitosis, and meiosis. Due to limitations of space and expertise of the author, this subject is not explored here in any depth, but we hope and intend that the

scale and nature of these challenges will be addressed in some detail in future contributions of authors and editors to *Frontiers in Plant Genetics and Genomics*.

THE EVOLUTIONARY PROCESS

Much has been learned in the first hundred years of genetics research about the molecular basis of mutations and about the processes of evolution. Tools and methods now exist that at least in theory allow us to determine the specific DNA changes that underlie essentially any mutation and even to compare the complete set of DNA variations that distinguish individuals, whether they are members of the same species or distantly related species. DNA sequencing technology continues to advance at an incredible rate, so it is obvious that geneticists and genomicists will have massive amounts of sequence information to compare individuals and species. Analyzing and making sense of all these data comprises the new field of comparative genomics and is a major challenge in its own right. However, the ultimate goal is to understand this diversity in terms of the mechanisms underlying the evolutionary process. This will require not only DNA sequence information, but also contributions from all the subspecialties of genetics and genomics – no doubt led by the fields of molecular evolution and population genetics, but with important contributions from the whole of genetics and genomics. Again, there is a set of major challenges that we hope will be explored in future contributions by authors and editors to *Frontiers in Plant Genetics and Genomics*.

EPIGENETICS – BETWEEN GENOTYPE AND PHENOTYPE

It was, of course, Waddington (1942a) who first proposed the terms “epigenetics” and “epigenotype” to describe the landscape between the gene and its final expression in the whole organism. In the modern era, these terms have become more focused on chromosomally based epigenetic information, rather the whole of physiological and developmental processes to which Waddington referred. Today we speak in terms of the “epigenome,” which is broadly defined to include all chromosomal modifications, including not only DNA modifications, but also the many modifications

of chromatin proteins and complexes. The epigenome determines both the expression of the genes and the inheritance of “epigenetic states,” mitotically and meiotically. Because many of these modifications appear to be “programmable” and to be “read out” to influence chromosomal functions, geneticists began to speak about 10 years ago of a “histone code” or “histone language,” but now of an “epigenomic code” or “epigenomic language” in order to encompass all chromosomal modifications, not only those of histones. Determining the language of the epigenome is clearly a major challenge for the twenty-first century. Also, we should be prepared for the likelihood that the epigenomic language of plants differs importantly from the epigenomic languages of fungi and animals: each kingdom possessing a distinct language, all descended from the “ur-language” of their common ancestor that existed a billion years ago.

FROM PHENOTYPE TO GENOTYPE – THE INHERITANCE OF ACQUIRED CHARACTERS?

The question of whether the “experiences” of an organism that induce adaptive somatic responses can be inherited have long been argued, but primarily as if the only explanations were the simple Lamarckian and Darwinian views, and when discussed it is often as if the question had been settled long ago. Nonetheless, it is still a live subject that represents a major challenge for biology, one that plant systems appear to be particularly well suited to address.

Waddington (1942b) was the first to propose an intermediate explanation of the inheritance of acquired characters between these two extremes. His main thesis was “that developmental reactions, *as they occur in organisms submitted to natural selection*, are in general canalized. That is to say, they are adjusted so as to bring about one definite end result regardless of minor variation in conditions during the course of the reaction” to selection. Waddington proposed that this “buffering,” which he preferred to call “canalization,” “ensures the production of the normal, that is, optimal type” of an organism “in the face of the unavoidable hazards of existence.” His application of the concept of canalization to “the inheritance

of acquired characters” was based on his interpretation that alternative paths of canalization of development occur when the environment plays an important role in the appearance of a new characteristic favored by natural selection.

Waddington’s proposed explanation for the inheritance of acquired characters’ basically comes down to the possibility that a mutation will sometimes arise that would favor an alternative, canalized path to the one normally expressed in the original genotype, and so “fix” that new path genetically. A key aspect of the proposal was that adaptation to the environmental stimulus precedes and is later superseded by a novel genetic alteration. Viewed from the perspective of the new genotype, the environmental influence on the original genotype would be said to “phenocopy” the new genotype in the original genotype.

In sum, Waddington’s thesis was that an organism may first adapt (physiologically or developmentally) to a selective force by switching between canalized paths, and that eventually a genetic mutation would also arise that favors the more adaptive path under the selective conditions, stabilizing it genetically and precluding the original path. Waddington’s “third explanation” was clearly Darwinian, not Lamarckian, because there was no influence of the environment on the occurrence of the specific mutation that would heritably stabilize (or “fix”) the proposed alternative path.

Several decades later Barbara McClintock, having observed the rapid, direct induction of new, heritable states by both environmental and developmental influences, proposed a middle explanation that truly encompasses the Darwinian and Lamarckian views and was not in conflict with Darwinism – despite some objections to the contrary; for further discussion, see Jorgensen (2004). In McClintock’s own words: “I believe there is little reason to question the presence of innate systems that are able to restructure a genome. It is now necessary to learn of these systems and to determine why many of them are quiescent and remain so over very long periods of time only to be triggered into action by forms of stress, *the consequences of which vary according to the nature of the challenge to be met* [emphasis added]” (McClintock, 1978).

McClintock (perhaps the greatest of many great maize geneticists) clearly stated the challenge in her Nobel lecture, challenging biologists “to determine the extent of knowledge the cell has of itself, and how it utilizes this knowledge in a ‘thoughtful’ manner when challenged” (McClintock, 1984). Understanding of the dynamic behaviors of the genome, the epigenome, the nucleus, and the cell has advanced tremendously in the decades since McClintock laid out this perhaps grandest of all the grand challenges of genetics, and so we are far better positioned now to determine the mechanisms

by which the cell and the organism utilize their “knowledge in a ‘thoughtful’ manner when challenged.”

REFERENCES

- Jorgensen, R. A. (2004). Restructuring the genome in response to adaptive challenge: McClintock’s bold conjecture revisited. *Cold Spring Harb. Symp. Quant. Biol.* 69, 349–354.
- McClintock, B. (1978). Mechanisms that rapidly reorganize the genome. *Stadler Genet. Symp.* 10, 25.
- McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792.
- Stadler, L. J. (1954). The gene. *Science* 120, 811–819.
- Waddington, C. H. (1942a). The epigenotype. *Endeavour* 1, 18–20.
- Waddington, C. H. (1942b). Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565.

Received: 25 November 2010; accepted: 20 December 2010; published online: 31 December 2010.

Citation: Jorgensen RA (2010) Of genes and genomes: challenges for the twenty-first century. *Front. Plant Sci.* 1:1. doi: 10.3389/fpls.2010.00001

This article was submitted to *Frontiers in Plant Genetics and Genomics*, a specialty of *Frontiers in Plant Science*.

Copyright © 2010 Jorgensen. This is an open-access article subject to an exclusive license agreement between the authors and the *Frontiers Research Foundation*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.