Evolutionary Biology

What is a gene? A two sided view

--Manuscript Draft--

Manuscript Number:	EVOL-D-16-00043R1
Full Title:	What is a gene? A two sided view
Article Type:	Essay
Keywords:	epigenetics; Evolution; genotype; phenotype; DNA
Corresponding Author:	Giorgio Camilloni Universita degli Studi di Roma La Sapienza Roma, ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universita degli Studi di Roma La Sapienza
Corresponding Author's Secondary Institution:	
First Author:	Adriano Bolondi
First Author Secondary Information:	
Order of Authors:	Adriano Bolondi
	Federica Caldarelli
	Francesca Di Felice
	Diletta Durano
	Giorgia Germani
	Luca Michetti
	Alessandra Tramutolo
	Gioacchino Micheli
	Giorgio Camilloni
Order of Authors Secondary Information:	
Funding Information:	Istituto Pasteur-Fondazione Cenci Bolognetti — Dr. Giorgio Camilloni
	Ministero dell'Istruzione, dell'Università e della Ricerca (Epigenomics Flagship Project EpiGen) — Dr. Giorgio Camilloni
Abstract:	
Response to Reviewers:	RESPONSES TO REVIEWER'S #2 COMMENTS (bold: reviewer's comments) •Figure 1: o "Outer elements": This is now replaced with "external elements", as suggested. A few changes have been been made to the legend of the figure. o DNA phenotype: We agree that histones contribute to the phenotype of the gene. On the other hand, in the figure and in the text our attempt is to concentrate more specifically on the phenotype of the DNA molecule. In the figure we already identify "DNA phenotype " as "molecular constitution and epigenetic modification of DNA". We feel that extending this definition as suggested by the reviewer to chromatin/chromosome would not accurately reflect the major focus of our essay. •English expression: The text has been reviewed and some formulations have been improved. •E.F.Keller reference: We have introduced the suggested citation.

	•Page 1, line 5: As suggested, "and the resultant of" has been replaced with "subject to". •Page 2, line 4: "Several years" is now “decades”. •Page 2, 2nd paragraph: The word “continuous” has been deleted to avoid possible misinterpretations with the concept of split genes. •Genotype vs phenotype section: o 2nd line: our intention was to refer to the two alleles in a single individual. The sentence has been now modified to reflect this. o Bottom of 2nd page, "phenotypic inheritance": This has been replaced with "gene phenotype inheritance", according to the reviewer’s suggestion. o End of section: -Coding essence: the last three lines have been slightly modified in order to acknowledge that the coding essence is related to the environment mainly "through natural selection". -The physical reality of DNA is common to all species: We would like to maintain our original formulation ("the physical reality of DNA as a molecule") since, in our opinion, it is sufficiently explicative to unambiguously identify the general structural/chemical invariance level which we want to address without entering into sequence details. •The epigenetic scenario: o Waddington reference: The year has been corrected. o Second paragraph: "genetic sequence" is now "gene", as suggested by the reviewer. o Third last line: -Epigenetics does act/could act: As suggested, the statement has been modified (and merged with the following one) to acknowledge that epigenetics "represents " rather "could represent ". -Variation of gene phonotype (examples): As suggested by the reviewer, in the revised version we briefly address examples of gene phonotype modifications. The corresponding references have been added. •The double faced gene: o Last line: Following the reviewer’s suggestion, the sentence has been modified to reflect that the view we propose also brings the phenotype and genotype domains closer together. o Commenting further on homology/homoplasy and development/evolution: These aspects are certainly very intriguing. However, we would rather not go deeper into them since we are convinced that their high complexity requires a more extensive approach which is beyond the focus of our paper and cannot be adequately dealt with in a few conclusive statements.

Bolondi A.[(1)]**, Caldarelli F.** [(1)]**, Di Felice F.** [(1)]**, Durano D.** [(1)]**, Germani G.** [(1)]**, Michetti L.** [(1)]**, Tramutolo A.** [(1)]**, Micheli G.** [(2)] **and Camilloni G.** [(1, 2, 3)]

What is a gene? A two-sided view.

[(1)] Dipartimento di Biologia e Biotecnologie, Sapienza, Università di Roma. Piazzale A. Moro 5. 00185 Roma, Italy

[(2)] Istituto di Biologia e Patologia Molecolari, CNR, Roma; Piazzale A. Moro 5. 00185 Roma, Italy

[(3)] Istituto Pasteur Italia - Fondazione Cenci Bolognetti e Dipartimento di Biologia e Biotecnologie - Sapienza -Università di Roma; Piazzale A. Moro 5. 00185 Roma, Italy

Corresponding author:

email: giorgio.camilloni@uniroma1.it

tel. +390649912808; fax +390649912500

Acknowledgements: This work was partially supported by the Istituto Pasteur-Fondazione Cenci Bolognetti (Università di Roma La Sapienza) and by the Epigenomics Flagship Project EpiGen (Italian Ministry of Education and Research, National Research Council)

The concept of gene: a brief history

Since its inception (Johannsen 1909) the notion of gene has evolved continuously (Keller 2000). In the past century the point of view about genes has been dominated by genetics up to the ‘30s. The great geneticists of that period, such as Sutton, Morgan, Bridges and Muller (Portin 2002) strongly contributed to the birth of that view. According to the studies on the transmission of characters, the gene is regarded as the indivisible unit of inheritance subject to mutations and genetic recombination. Successively, during the '40s, the studies performed on *Neurospora crassa* allowed to link the concept of gene to the synthesis of a given enzyme, yielding the well-known "one gene - one enzyme" theory by Beadle and Tatum (Beadle and Tatum 1941). This genetic-biochemical conception had its turning point towards a molecular view in 1953, when Watson and Crick solved the basic structure of DNA (Watson and Crick 1953). The association of DNA with the genetic material had been made ten years before by Avery and coworkers (Avery et al. 1944) and had been confirmed by Hershey and Chase in 1952 (Hershey and Chase 1952). The notion that the physical base of inheritance resides in DNA paved the way to several important findings, such as those of Jacob and Monod (Jacob and Monod 1961), and became widely accepted over the '70s. This led to amend the concept of gene, identifying it with a continuous DNA sequence responsible for the synthesis of a given mRNA and consequently of a polypeptide.

The co-linearity involving gene, RNA and protein has been the standard model until the middle of the '70s. Then, following the introduction of recombinant DNA technologies and the consequent cascade of new discoveries, ideas concerning the essence of a gene did no longer fully correlate with experimental data. The concept of gene evolved further as the limits of the previous view became evident in the light of major novel acquisitions: i) the discovery of repeated genes that do not code for proteins (Davidson and Britten 1973); ii) the discovery of split genes, which also demonstrated the absence of an absolute gene-protein co-linearity and the existence of a surplus of genetic material (Chow et al. 1977); iii) the capability of the cell to process RNA not just in a single way but in many different ways through alternative splicing, resulting in different proteins from the

same coding sequence (Horowitz et al. 1978). Moreover, in the '70s transposable elements, DNA sequences able to move within a genome and among genomes (Cohen 1976), were characterized at the molecular level, although their genetic identity had been predicted by the pioneering studies of Barbara McClintock decades before (McClintock 1947). These findings followed by several others, e.g. the discovery of overlapping genes (Normark et al. 1983), revealed an unexpected plasticity of the genome and the gene-protein co-linearity, with its organization able to change only by mutation or DNA recombination, became old-school.

More recently the development of sequencing technologies such as Next Generation Sequencing (NGS), which allows quick sequencing of whole genomes (Schuster 2008), allowed to observe that the entire mammalian genome is transcribed in a pervasive manner (Jensen et al. 2013). This implies that not only putative coding regions undergo transcription but that almost the entire genome is transcriptionally active and that the boundaries separating genes are not clearly delineated. This challenges the currently dominant concept of gene as a unit delimited by specific initiation and termination points. Moreover, the advent of epigenetics and novel acquisitions on the tridimensional structure of DNA and of its supramolecular complexes, provide experimental support to the notion that, in terms of hereditability, information can be associated also to different elements beyond a simple linear array of nucleotides. Present views see a gene as a DNA sequence converted into RNA through transcription. This RNA, in turn, may or may not be translated into protein(s). The final products of transcription and/or translation specify a character *sensu stricto*. This implies that the function of a given RNA or of a given protein can affect cell physiology at whatever level, both in a structural respect and/or by regulating or controlling other genes. This definition takes account of protein coding genes, non coding RNAs, split genes, alternative splicing, microRNAs and long non coding RNAs.

Genotype vs phenotype?

The term *genotype* refers to the entire set of genes in a cell, an organism, or an individual. In the diploid complement of an individual a gene for a particular character or trait may be present in two allelic forms. The notion of p*henotype*, on the other hand, identifies observable physical or biochemical features of an organism, as determined by both genetic makeup and environmental influences. Considering these definitions, the genotype is constituted by the DNA sequences representing genes. However, DNA has intrinsically also a phenotypic nature because most of its biochemical features are observable and measurable. According to this view all DNA structures, intended as simple arrays of pentose sugars linked by phosphodiester bonds and carrying nitrogen bases bound by N-glycosidic linkages, represent the phenotypic essence of a gene (we can call this the *DNA phenotype* to distinguish it from the *phenotype at the organism*, fig. 1) while its genotypic nature resides in the coding potential underlying transcription. In this morpho-functional perspective a gene is envisaged as a DNA tract endowed with phenotype and genotype at the same time (coding regions), while other regions act as pure phenotype (non coding regions). The genotype nature of a gene could not exist without the supporting framework provided by the DNA phenotype. At the organism level the genotype accounts for the observed phenotype (fig.1).

Traditionally, when discussing about characters associated to DNA coding properties, the phenotype nature of the molecule has been disregarded. From an evolutionary standpoint the information flux that characterizes the present reality is the result of natural selection processes that allowed to fix defined genetic and morphologic traits in response to the surrounding environment and its variations. In essence, we know what genetic inheritance is all about, i.e. the transmission of the genetic complement of a generation to its offspring. However, DNA has also its physical nature as an ensemble of atoms occupying a given space and leading to the formation of a defined structure. Beyond the informational-genotypic component it is necessary to consider also the structural-phenotypic aspect of DNA. These two natures mutually affect each other and the boundary separating them is often elusive. Banking on these considerations, the concept of *gene phenotype inheritance* can be introduced, based on the structure of the DNA molecule and on its

ability to be conserved and transmitted in all living organisms since 3.5 billions of years: while the coding essence is related to the environment, mainly through natural selection, and is subject to large variations, the physical reality of DNA as a molecule remains the same during evolution and is common to all living species.

The epigenetic scenario

Since 1942 (Waddington 1942) it has been proposed that, at least for certain characters, a further layer controlling inheritance exists, acting on top of the genetic level: epigenetics. The transmission of the functional state (epigene) is as important as the function encoded in a given gene. The functional state can be affected by physical modifications of a gene (e.g. methylation) (Schübeler 2015) or carried by other chromatin components (i.e. modifications of histone tails; nucleosome positioning and/or occupancy) (Kouzarides 2007). These modifications may directly depend on the environment (Feil and Fraga 2012). Some phenotypic modifications borne by DNA may pass also to offspring, although in minimal part (Blaze and Roth 2015). Hence, the genotype can acquire novel informational content from the environment without changing its sequence but simply by modifying its phenotype.

In this perspective epigenetics provides a clear link between the genotypic and the phenotypic nature of a gene. In other words, an epigenetic modification can actually affect the functional status of a gene determining its *phenotypic* change through the alteration of the chemical structure of the nucleic acid, but at the same time it can also determine a *genotypic* regulation by promoting or silencing gene expression and/or by acting on the coding properties of the nucleic acid. Some epigenetic modifications are inheritable, as is the case for DNA methylation, which in turn can trigger additional epigenetic alterations. Epigenetic modifications are also able to alter the genotype by modifying the phenotype. An exemplification comes from experiments showing that in rodents the DNA methylation profiles of germ cells may be altered, with consequent strong transgenerational potential, following exposure of individuals to chemical stress conditions during

adult or prenatal life (Pacchierotti and Spanò 2015). In general, there is now substantial agreement that environmental and stress factors are strictly related to epigenetic modifications affecting gene expression (Roth 2013). In essence, epigenetics as transducer of external signals represents an additional variability source, acting as an additional evolutionary driving force together with natural selection and genetic drift (Schrey et al. 2012).

The double faced gene.

The need to account for all currently available experimental observations has reshaped the concept of gene, turning an essentially mechanistic unit predominant during the '70s into a quite abstract, open and generalized entity whose contour appears less defined as compared to the past. The more data are gathered, the greater an abstraction effort is required in order to understand and define the essence of a gene. The boundaries of the object we are investigating (the gene itself) become fuzzy as we move closer. The same happens at the scale of an electron, with its double nature of particle and wave, where measurements become probabilistic and not absolute. Can also the essence of the gene be considered double-faced? In this respect genotypic and phenotypic entities of a gene would coexist and mix reciprocally (fig. 1). This harmonizes present knowledge with current definitions and predisposes for remodeling of our thinking as a consequence of future discoveries. This two-sided view of the gene brings its phenotype and genotype domains closer together and allows to combine the genetic and epigenetic aspects in a unique solution, being structural and functional at the same time and simultaneously able to include the different levels in an overlapping *unicum*.

FIGURE LEGEND

The two-faced nature of genes. The DNA phenotype, i.e the molecular constitution of the DNA and its epigenetic modifications, determines the coding information, i.e. the genotype. Through gene expression the organismal phenotype is obtained. Following outer inputs the latter may, in turn, influence the DNA phenotype. The double-faced image refers to the ancient roman myth of Janus bifrons.

REFERENCES

Avery, O.T., MacLeod, C.M., McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III. J. Exp. Med., 79, 137-158.

Beadle, G.W., Tatum, E.L. (1941). Genetic control of biochemical reactions in Neurospora. Proceedings of the National Academy of Science USA, 27, 499-506.

Blaze, J., Roth, T.L. (2015). Evidence from clinical and animal model studies of the long-term and transgenerational impact of stress on DNA methylation. Semin Cell Dev Biol., 43, 76-84.

Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell, 12, 1-8.

Cohen, S.N. (1976). Transposable genetic elements and plasmid evolution. Nature, 263, 731-8.

Davidson, E.H., Britten, R.J. (1973). Organization, transcription, and regulation in the animal genome. Q Rev Biol., 48, 565-613.

Feil, R., Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet., 13, 97-109.

Hershey, A.D., Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. Journal of General Physiology, 36, 39-56.

Horowitz, M., Bratosin, S., Aloni, Y. (1978). Polyoma infected cells contain at least three spliced late RNAs. Nucleic Acids Res., 5, 4663-75.

Jacob, F., Monod, J. (1961). Molecular and biological characterization of messenger RNA. Journal of Molecular Biology, 3, 318-356.

Jensen, T.H., Jacquier, A., Libri, D. (2013. Dealing with pervasive transcription. Mol Cell., 52, 473-84.

Johannsen, W. (1909). Elemente der exakten erblichkeitslehre (pp. 143-144). Gustav Fischer, Jena.

Keller, E.F. (200). The Century of the Gene. Harvard University Press, Cambridge (MA).

Kouzarides, T. (2007). Chromatin modifications and their function. Cell, 128, 693-705.

McClintock, B. (1947). Cytogenetic studies of maize and Neurospora. Carnegie Institute Washington Yearbook, 46, 146-152.

Normark, S., Bergström, S., Edlund, T., Grundström, T., Jaurin, B., Lindberg, F.P., Olsson, O. (1983). Overlapping genes. Annu Rev Genet., 17, 499-525.

Pacchierotti, F., Spanò, M. (2015). Environmental Impact on DNA Methylation in the Germline: State of the Art and Gaps of Knowledge. BioMed Research International, 2015, Article ID 123484

Portin, P. (2002). Historical development of the concept of the gene. J Med Philos., 27, 257-86.

Roth, T.L. (2013). Epigenetic mechanisms in the development of behavior: advances, challenges, and future promises of a new field. Dev Psychopathol., 25, 1279–1291.

Schrey, A.W., Coon, C.A., Grispo, M.T., Awad, M., Imboma, T., McCoy, E.D., Mushinsky, H.R., Richards, C.L., Martin, L.B. (2012). Epigenetic Variation May Compensate for Decreased Genetic Variation with Introductions: A Case Study Using House Sparrows (Passer domesticus) on Two Continents. Genet Res Int., 2012, articleID 979751.

Schübeler, D. (2015). Function and information content of DNA methylation. Nature, 517, 321-6.

Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. Nature Methods, 5, 16-8.

Waddington, C.H. (1942). The epigenotype. Endeavor, 1, 18-20.

Watson, J.D., Crick, F.H.C. (1953). A structure for deoxyribonucleic acid. Nature, 171, 737-738.