



HAL
open science

Pseudoalleles and gene complexes: the search for the elusive link between genome structure and gene function

Michel Morange

► To cite this version:

Michel Morange. Pseudoalleles and gene complexes: the search for the elusive link between genome structure and gene function. *Perspectives in Biology and Medicine*, 2016, 58 (2), pp.196-204. 10.1353/pbm.2015.0027 . hal-01347725

HAL Id: hal-01347725

<https://hal.science/hal-01347725>

Submitted on 21 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Pseudoalleles and gene complexes: the search for the elusive link between genome structure and gene function

Michel Morange, Centre Cavallès, République des savoirs:
Lettres, sciences, philosophie USR3608, Ecole normale
supérieure, 29 rue d'Ulm, 75230 Paris Cedex 05, France

E-mail: michel.morange@ens.fr

ABSTRACT

After their discovery in the first decades of the XXth century, pseudoalleles generated much interest among geneticists: they apparently violated the conception of the genome as a collection of independent genes elaborated by Thomas Morgan's group.

Their history is rich, complex, and deserves more than one short contribution. I will focus on two issues: the way the phenomenon of pseudoallelism suggested that the genome was more than a simple addition of independent genes, and the connection established between the formation of pseudoalleles during evolution and their functional roles. I will successively consider the first explanations for the origin of pseudoalleles elaborated in the mid-1930s, the metabolic/developmental sequential model proposed by Ed Lewis in the 1950s, the disappointments encountered with the *T*-complex in the 1970s, and the fading of the previous models after the molecular characterization of the pseudoallelic gene complexes in the 1980s.

Genomes are more than collections of genes, but their structures are the result of a complex evolutionary history that leaves no place for simplistic models.

Pseudoalleles are closely linked genes that have similar functions. Their proximity on the chromosome makes their distinction by the complementation tests traditionally used by geneticists difficult. For this reason, and because they have similar functions, they were initially often considered as allelic forms of the same gene, hence their name. The *Hox* cluster is an emblematic example of a pseudoallelic gene complex. The history of research on pseudoalleles is rich and complex. The first observations were made very early, but remained puzzling until a simple model explaining their formation and characteristics emerged in the middle of the 1930s: pseudoalleles originated by gene duplication, the two copies of the gene remaining closely associated on the chromosome, but progressively diverging in structure and function. This model did not prevent the active discussion of new observations on pseudoalleles in the following years. There is an additional, more important reason for a historian to be interested in this system. The study of pseudoalleles was an unsuccessful attempt to bridge the gap between genes and genomes, and to find in the structural organization of the genome clues to how genes function. As Edward B. Lewis put it in 1955: “The phenomenon of pseudoallelism promises to contribute much to our understanding of the gene – how it functions, how it mutates, how it evolves” (Lewis 1955). The history of pseudoallelism illustrates the difficulty of demonstrating that genomes are more than the addition of individual genes.

Research on pseudoalleles has characteristics that are worth underlining at the outset. There was a pre-eminence of a limited number of experimental systems (*Bar* (an ensemble of mutations affecting the shape of the eyes), *achaete-scute*, *bithorax* (a part of the *Hox* cluster) in *Drosophila*, the *T-*

complex in mice) that attracted the most eminent geneticists (for *Bar*, Alfred Sturtevant, Theodosius Dobzhansky, Calvin Bridges, Hermann Muller, Sewall Wright, A. A. Prokofyeva-Bolgovskaya, K. V. Kossikov) over a long period of time. The study of pseudoalleles was a dual issue, concerning both functional and evolutionary biologists, the mechanism of their formation being linked with the functional role they played in the organism. This is probably why so many different issues—position effects (*i.e.* the fact that the activity of genes may change according to their position on the chromosomes), the mechanisms of gene duplication, the role of chromatin and the regulation of gene expression—were, transiently or not, associated with their study.

I will successively discuss four episodes in this complex history: the early observations on pseudoalleles and gene complexes, and the mechanistic models provided to explain their formation; the metabolic/developmental sequential model proposed by Ed Lewis in the 1950s; the long and tumultuous history of the *T*-complex (1927-1980); and the “golden age” of gene complexes (1980-1990) with the precise description of their structure at the molecular level. A full history of pseudoalleles would require more than a book!

I. The early observations on pseudoalleles

To understand why the existence and functions of pseudoalleles have always been problematic, it is necessary to bear in mind the vision of the genome held by Thomas Morgan and by most American geneticists: “So far as we can judge from the action of mutated genes, the kind of effect produced has as a rule no relation to location of the gene in the chromosome. A gene may produce its chief effect on the eye colour, while one nearby may affect the gene structure, and a third, in the same region, the fertility of the male or of the

female. Moreover, genes in different chromosomes may produce almost identical effects on the same organs. One may say, then, that the position of the genes in the hereditary material is inconsequential in relation to the effects that they produce” (Morgan 1934). Given this view, it is quite remarkable that the early observations on the existence of pseudoalleles were made by collaborators of Morgan, Alfred Sturtevant and Hermann Muller, as well as by the Russian geneticist A. S. Serebrovsky. A mechanism for their formation and extension by unequal crossing over was rapidly proposed. This was confirmed by Calvin Bridges in 1936 through the direct observation of these pseudoallelic complexes on the giant chromosomes of *Drosophila* (Bridges 1936), and simultaneously by Hermann J. Muller in a Russian publication. This phenomenon, occurring at the gene level, had clear analogies with other phenomena described at higher levels of organization in animals: repetition of parts of the organism was a well-known phenomenon revealed by comparative anatomy, and used by evolutionary biologists to explain some of the evolutionary transformations observed in organisms. The origin of life was identified with the formation of the first genes by Muller: therefore, it was not surprising to obtain evidence that new genes originated from pre-existing ones, as every cell comes from the division of a pre-existing cell, and life always originate from life (Muller 1936). In this evolutionary conception, the formation of pseudoalleles was seen only as a transient step, a hypothesis in contradiction with the observation that at least some of these gene complexes seemed to have been conserved in evolution. The existence of pseudoalleles also received attention from the rare geneticists who opposed the particulate conception of the gene, such as Richard Goldschmidt. For them, the existence of pseudoalleles was a strong argument against the “theory of the

gene” (Goldschmidt 1937).

II. The metabolic/developmental explanation of the functions of pseudoalleles

In 1945, Norman Horowitz proposed an evolutionary scenario for the formation of long metabolic pathways (Horowitz 1945). His model was a consequence of the experiments of George Beadle and Edward Tatum who had confirmed in *Neurospora* a relation between genes and enzymes that was implicit since Cuénot and the early steps of genetics (Beadle, and Tatum 1941), but also of the model for the origin of life proposed by Alexander Oparin (Oparin 1938). He suggested that metabolic pathways had grown during evolution in a stepwise process from the end product to the substrate of the first enzyme by the progressive recruitment of mutations generating new enzymes catalyzing each of the steps of the metabolic pathway. This article was important because it was the first to address the issue of the evolutionary origin of the stepwise pathways described in metabolism.

The main contributor to the metabolic interpretation of the functions of pseudoalleles was Ed Lewis. He had acquired considerable experience of pseudoalleles, which he had been studying for more than ten years (Lewis 1941 and 1952). He proposed the sequential reaction model. Since an enzyme catalyzing a reversible reaction recognizes both the substrate and the product of the reaction, a simple mutation was able to orient the conversion of the product into another product, elongating the metabolic chains. After duplication, the different genes present in a pseudoallelic complex diverged to generate these long metabolic pathways (Lewis 1951). Lewis adapted this model to the pseudoallelic *bithorax* complex. This pseudoallelic complex was involved in the development of *Drosophila*. By transposing a hypothesis on the origin of

metabolic pathways to developmental pathways, Ed Lewis established a link between embryological mechanisms and the evolution of genomes. It was not a big leap, since developmental steps were often considered to be identical to metabolic steps.

The metabolic model also explained why pseudoallelic complexes are not transient, but are stable during evolution. Since the products of the genes (enzymes?) were probably physically very close to the genes themselves (a hypothesis that still remained acceptable for eukaryotes at that time), keeping the genes in close proximity meant keeping nearby the enzymes produced by these genes, an obvious kinetic advantage in a metabolic pathway since it limits the time of diffusion of the substrates and products. By extension, the physical association between the different genes involved in the same developmental pathway was also considered as beneficial. Ed Lewis also distinguished position pseudoallelism from other forms of position effects effective over long distances (variegation).

At the end of the 1950s, it was discovered that genes involved in the same metabolic pathway were often grouped on the bacterial genome in a structure later called operon, and transcribed into a unique RNA species. The existence of operons seemed to support the previous metabolic model of pseudoallelism (Lewis 1963), although nothing demonstrated that the genes present in an operon had a common evolutionary origin, and further experiments even demonstrated the opposite. At the end of the 1960s, many debates took place on the existence (or not) of operons in eukaryotes. Their absence – with few exceptions – became evident only in the 1980s when genetic engineering tools permitted the isolation and characterization of genes and mRNAs, definitively breaking the relation between operons

and pseudoalleles.

III. The *T*-complex

In the words of Virginia Papaioannou (1999), the *T*-complex educated a generation of developmental biologists (Papaioannou 1999). Mutations in the genes of this pseudoallelic complex were described as early as 1927 (Dobrovolskaia-Zavadskaia 1927), and the list of geneticists devoting time to the study of this complex included Leslie Dunn, Salomé Gluecksohn-Schoenheimer, Ernst Caspari, Mary Lyon, Dorothea Bennett and many others.

The phenotypic characteristics of the mutations were complex. The *T*-mutation was responsible for a short tail at the heterozygous state, and was lethal at the homozygous state. The different *t*-mutations were lethal for homozygotes, but heterozygotes were indistinguishable from wild-type animals (Bennett 1964). However, a *t*-mutation combined with a *T*-mutation generated tailless mice, which constituted a simple way to isolate new *t*-mutations. In addition, there was a biased transmission of the *t*-mutations that was interpreted as the effect of the expression of the products of the *T*-complex on the membranes of spermatozoa. A battery of different *t*-mutations was progressively described, interrupting early embryological development at different successive steps. In 1975, Dorothea Bennett proposed a model that described the role of the *T*-complex in development and explained the characteristics of the different *t*-mutations (Bennett 1975). The *T*-complex encoded a family of membrane proteins: each of them controlled one early step in embryogenesis. These membrane proteins generated new interactions between embryonic cells that transformed the characteristics of the cells and led to the expression of a different gene of the *T*-complex, allowing new intercellular interactions. The process

was repeated as many times as there were different genes in the *T*-complex.

There were various reasons for proposing that the *T*-complex encoded membrane proteins, but some of them came from a collaboration between a student of Dorothea Bennett, Karen Artz, and the laboratory of François Jacob, who had recently abandoned the study of bacteria and bacteriophages for that of mouse development (Morange 2000). The experimental approach was to raise specific antisera against embryonal carcinoma cells. These were stem cells of tumours of the gonads called teratocarcinomas, and they were considered to have characteristics in common with early embryonic cells. One of these antisera bound to a membrane protein of the embryo called the F9-antigen, from the name of the embryonal carcinoma cell line used to obtain it. It prevented the compaction of the embryo, a major step of development preceding the formation of a blastocyst, and the first differentiation event. It suggested that the F9-antigen might be involved in this process. By using complex immunocompetition experiments, Karen Artz demonstrated that the expression of the F9-antigen was reduced in embryos bearing the t^{I2} -mutation, the mutation of the *T*-complex that had the earliest phenotypic effect on development (Artzt, Bennett, and Jacob 1974 ; Artzt, and Bennett 1975).

Additional experiments performed by Karen Artz suggested that the F9-antigen contained a chain of β 2-microglobulin, a protein associated with the H2 immunocompatibility complex of the mouse, the equivalent of the HLA system of humans. The H2- and *T*- complex might have originated by duplication of a unique ancestral complex involved in the control of early embryogenesis. Even before these results, Walter Bodmer had elaborated an evolutionary scenario in which the genes of the H2/HLA complex were a by-product of the duplication of

genes involved in the control of differentiation (Bodmer 1972).

This functional model of the *T*-complex combined a precise mechanistic model of development and an evolutionary explanation of the way this genetic system (and development) progressively became more and more complex. Somehow it fulfilled the dreams that had been supporting the study of pseudoalleles since their discovery: to provide simultaneously an evolutionary and mechanistic explanation of their functions.

Unfortunately, the observations on which this model had been based were wrong: the expression of the F9-antigen was not altered in the t^{I2} -mutation, and β 2-microglobulin is not associated with it. But worst of all was the discovery that the *T*-complex did not exist! The different mutations thought to be in the same gene complex were scattered on the chromosome. Their apparent proximity was the artefactual result of a past chromosomal inversion that prevented genetic recombination and generated deformed genetic maps. These negative results sounded the death knell of the excessive hopes that had been invested in the pseudoallelic gene complexes. Pseudoalleles were not the keys that would open the door to an understanding of genome structure.

IV. The golden age of pseudoallelic gene complexes (1980-1990)

This decade corresponded to the molecular characterization of pseudoallelic gene complexes, made possible by the newly introduced tools of genetic engineering. Particular attention was focused on the HLA and H2 gene complexes (Klein, Figueroa, and Nagy 1983), the globin gene complex (Collins, and Weissman 1984), the mouse and human *Hox* gene complex (homologous to the *bithorax* and *antennapedia* gene

complexes of *Drosophila*) (Duboule and Dollé 1989), and more recently the odorant receptor gene complex (Kratz, Dugas, and Ngai 2002).

The situation was shown to be different for each complex. In the HLA and H2-complexes, genes such as those encoding some of the components of the complement (a complex of proteins involved in the destruction of infected cells), structurally unrelated to the major histocompatibility genes, had been associated with the complex during evolution for functional reasons. In most cases, the existence of a complex was the result of a process of gene duplication, the importance of which was reemphasized by Susumu Ohno in 1970 (Ohno 1970). The different genes present in the complex had often acquired different functions and regulation. The existence of a complex was clearly the consequence of the mechanism of generation of new genes; but the genes having become independent, the future maintenance of a complex was problematic.

In some cases, the close association of the genes and their position on the chromosome remained important for their correct expression in the time and space of the organism. Such is the case of the globin gene complex, and even more significantly of the famous *Hox* gene complex. The position of a gene in the complex is related to its timing and level of expression along the anterior-posterior axis of the embryo (as well as in the limbs); and the structural organization of the complex has been conserved “from *Drosophila* to humans” (Duboule, and Dollé 1989).

Gene complexes are a by-product of an important phenomenon for the evolution of genomes – gene duplication – but also of the complex mechanisms of gene regulation that were progressively brought to light. The most significant step for a reappraisal of the physiological significance of the

existence of gene complexes was the discovery of enhancers by Walter Schaffner in 1981 (Banerji, Rusconi, and Schaffner 1981; Morange 2014). These DNA sequences activate gene expression at a long distance, in one or other orientation, in the 5' or 3' position relative to the gene they regulate. These enhancer sequences are formed of binding sites for transcription factors. Activation of transcription requires the direct interaction of enhancers with promoters through the formation of loops.

Recently, thanks to a huge accumulation of data on gene regulation as well as to the development of specific programs such as ENCODE, which is designed to provide a full description of regulatory elements in a genome, a new vision of enhancer action has emerged (De Laat, and Duboule 2013). Enhancers are highly abundant in a genome like the human genome, 20 to 30 times more abundant than the genes themselves. According to Chromosome Conformation Capture (3C) studies, enhancer – gene contacts might be predominantly restricted to topologically associated domains (TADs), coherent chromatin domains (or blocks) along the linear genomic DNA.

The existence of gene complexes is the result of four unrelated and independent phenomena and processes: duplication of genes, their progressive functional divergence, the conservation (or not) of enhancers during evolution, and the organization of chromatin in different TADs. The existence (or inexistence) of a complex and its functional integration are the results of the relative weight of these different processes.

Conclusion

Thomas Morgan was not utterly right. The functions of genes can depend on their position in the genome, and on their integration in gene complexes. Gene duplication is an

important mechanism of genome evolution. But acknowledging the existence of these phenomena only yields ambiguous clues to the relation between genome organization and gene function.

The difficulty of apprehending the complex relations between genome structure and gene function is another example of the difficulty of dovetailing evolutionary (historical) and functional explanations in biology. Whereas functional approaches would favour simple solutions, the long evolutionary history of organisms clouds the picture and prevents any generalization.

Acknowledgements:

I am indebted to Ute Deichmann, David Marsh and the anonymous reviewer for critical reading of the manuscript.

References

- Artzt, Karen, Dorothea Bennett, and François Jacob. 1974. "Primitive teratocarcinoma cells express a differentiation antigen specified by a gene at the *T*-locus in the mouse." *Proc. Natl. Acad. Sci. USA* 71: 811-4.
- Artzt, Karen, and Dorothea Bennett. 1975. "Analogies between embryonic (T/t) antigens and adult major histocompatibility (H-2) antigens." *Nature* 256: 545-7.
- Banerji, J., S. Rusconi, and W. Schaffner. 1981. "Expression of a β -globin gene is enhanced by remote SV40 DNA sequences." *Cell* 27: 299-308.
- Beadle, George W, and Edward L. Tatum. 1941. "Genetic control of biochemical reactions in *Neurospora*." *Proc. Natl. Acad. Sci. USA* 27: 499-506.
- Bennett, Dorothea. 1964. "Abnormalities associated with a chromosome region in the mouse." *Science* 144: 260-7.
- Bennett, Dorothea. 1975. "The T-locus of the mouse." *Cell* 6:

441-54.

Bodmer, W. F. 1972. "Evolutionary significance of the HL-A system." *Nature* 237: 139-45.

Bridges, Calvin B. 1936. "The "BAR" gene a duplication." *Science* 83: 210-1.

Collins, F. S., and S. M. Weissman. 1984. "The molecular genetics of human hemoglobin." *Prog. Nucl. Acid Res. Mol. Biol.* 31: 315-482.

De Laat, Wouter, and Denis Duboule. 2013. "Topology of mammalian developmental enhancers and their regulatory landscapes." *Nature* 502: 499-506.

Dobrovolskaia-Zavadskaia, N. 1927. "Sur la modification spontanée de la queue chez la souris nouveau-née et sur l'existence d'un caractère (facteur) héréditaire "non-viable"." *C. R. Soc. Biol.* 97: 114-6.

Duboule, Denis, and Pascal Dollé. 1989. "The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes." *The EMBO J.* 8(5): 1497-505.

Goldschmidt, Richard. 1937. "Spontaneous chromatin rearrangements and the theory of the gene." *Proc. Natl. Acad. Sci. USA* 23: 621-3.

Horowitz, N. H. 1945. "On the evolution of biochemical syntheses." *Proc. Natl. Acad. Sci. USA* 31: 153-7.

Klein, J., F. Figueroa, and Z. A. Nagy. 1983. "Genetics of the major histocompatibility complex: the final act." *Annu. Rev. Immunol.* 1: 119-42.

Kratz, Erica, Jason C. Dugas, and John Ngai. 2002. "Odorant receptor gene regulation: implications from genomic organization." *Trends Genet.* 18(1): 29-34.

Lewis, E. B. 1941. "Another case of unequal crossing-over in *Drosophila melanogaster*." *Proc. Natl. Acad. Sci. USA* 27: 31-5.

- Lewis, E. B. 1951. "Pseudoallelism and gene evolution." *Cold Spring Harbor Symp. Quant. Biol.* 16: 159-74.
- Lewis, E. B. 1952. "The pseudoallelism of *white* and *apricot* in *Drosophila melanogaster*." *Proc. Natl. Acad. Sci. USA* 38: 953-61.
- Lewis, E. B. 1955. "Some aspects of position pseudoallelism." *Am. Nat.* 89: 73-89.
- Lewis, E. B. 1963. "Genes and developmental pathways." *Am. Zool.* 3: 33-56.
- Morange, Michel. 2000. "François Jacob's lab in the seventies: the *T*-complex and the mouse developmental genetic program." *Hist. Phil. Life Sci.* 22: 397-411.
- Morange, Michel. 2014. "Enhancers: their existence and characteristics have raised puzzling issues since their discovery". *J. Biosci.* 39: 741-5.
- Morgan, Thomas. 1934. "Nobel lecture."
- Muller, Hermann J. 1936. "Bar duplication." *Science* 83: 528-30.
- Ohno, Susumu. 1970. *Evolution by gene duplication*. New York: Springer.
- Oparin, Alexander I. 1938. *The origin of Life*. New York: MacMillan.
- Papaioannou, Virginia E. 1999. "The ascendancy of developmental genetics, or how the T complex educated a generation of developmental biologists." *Genetics* 151: 421-5.