

Roots, strategies and prospects of functional genomics*

Werner Arber

Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

This essay traces the historical development of classical and molecular genetics from their early roots to the actual research strategies and their prospects. Attention is also given to the risk evaluation of genetic engineering by comparing designed genetic alterations with the spontaneous genetic variation known to form the substrate for biological evolution.

CLASSICAL genetics has its roots in the 19th century, when Gregor Mendel carried out experiments with peas displaying phenotypically distinct traits which got inherited to the progeny. When different variants were sexually crossed, recombinants appeared with reproducible frequencies and these experiments led to the establishment of linear maps of different traits. In the early decades of the 20th century cytology succeeded in certain cases to relate specific morphological alterations in the chromosomes with the variation of particular phenotypical traits. The term gene was then introduced for the chromosomal component responsible for specific phenotypes. However, the chemical nature of the gene had remained unknown.

This situation changed when microbial genetics involving both bacteria and bacterial viruses took its origin in the 1940s. These microorganisms could be shown to undergo mutation like higher organisms. This enabled the investigators to demonstrate that determinants of specific phenotypic traits could be transferred from a donor to a recipient bacterium. Within a few years, three distinct transfer mechanisms could be detected: transformation¹, conjugation² and phage-mediated transduction³.

Chromosomes had been shown to be composed of nucleic acids and proteins. A careful chemical separation of these components followed by transformation experiments could then clearly show that it is the DNA which confers on the recipient bacteria the ability to develop the specific phenotypical traits displayed by the donor bacteria¹, and such acquired traits were inherited to the progeny. Thus, DNA was identified as the carrier of genetic information.

Just a few years later, in 1953, Watson and Crick published the double-helical structure of DNA⁴. It then

became obvious that genetic information was contained in the linear sequences of nucleotides forming the building blocks of the very long, filamentous DNA molecules. The double-stranded nature of DNA molecules was postulated to ensure a high degree of fidelity in the replication of these molecules.

At that time it was known that proteins must represent the major products of genes and that proteins are built up by linear sequences of amino acids. Within about 15 years the mechanism of protein synthesis was unravelled, and so was the genetic code which serves in all living organisms to translate the genetic message contained in the DNA into the protein language. Hence, at the end of the 1960s the direct correlation between linear sequences of nucleotides in nucleic acids and the linear sequences of amino acids in proteins was clearly established. However, efficient methods to experimentally determine large extents of such sequences were still missing. It was also clear that genes consisted of (a) open reading frames serving as instruction on how to synthesize the proteins as gene products, and (b) specific nucleotide sequences serving to control the time and efficiency by which specific genes should become expressed.

In the meantime, microbial genetics had revealed how some small, autonomously replicating DNA molecules – plasmids and viral genomes – could become involved in horizontal gene transfer upon conjugation and specialized phage-mediated transduction. Interestingly, hybrid DNA molecules were naturally formed in these processes, and they were composed of the replicative machinery of the plasmid or viral genome on the one hand, and of a segment of the bacterial host chromosome on the other. The investigations of these phenomena largely benefited from an urgent medical interest residing in the rapid spreading of antibiotic resistances of bacteria as a consequence of the wide use of antibiotics. An interesting conclusion from these studies was that horizontal gene transfer, which could span wide evolutionary distances, was fostered by natural gene vectors (plasmids and viruses) and that these phenomena played an important role in the generation of genetic variations forming the substrate for natural selection in biological evolution.

However, in the 1950s and 1960s it also became clear that bacteria succeed by a number of different strategies to hold the frequency of acquisition of foreign genetic information low. One of these strategies is the widely

*Based on the Rajiv Gandhi Science and Technology Lecture entitled 'Molecular Genetics: Roots, Strategies and Applications' delivered on 4 December 1998 in Chandigarh.
e-mail: Werner.Arber@unibas.ch

encountered phenomenon of restriction and modification of DNA⁵. Restriction–modification systems allow bacteria to specifically distinguish foreign DNA from the cell's own DNA. As a consequence, foreign DNA becomes cleaved into fragments upon its entry into the cell. DNA fragments with open ends are recombinogenic, which favours their integration into the host genome; but they usually become rapidly degraded by exonucleases. This situation ensures that the acquisition of foreign genetic information by bacteria is inefficient and occurs usually in small steps, i.e. by taking up relatively little foreign genetic information per event. This makes much sense as an evolutionary strategy⁶. By the end of the 1960s, restriction endonucleases were purified and could reveal their cleavage of unmodified DNA molecules *in vitro*⁷.

On the basis of knowledge acquired by research on microbial genetics, around 1970 investigators succeeded in preparing recombinant DNA molecules *in vitro* and to get them replicated after their transfer into appropriate host cells⁸. In these experiments plasmids and viral genomes served as gene vectors into which fragments of DNA, often prepared by restriction cleavage, were spliced. The resulting hybrids were introduced into host bacteria either by transformation or by infection after *in vitro* packaging into virus particles. In the host cells the hybrid DNA molecules became replicated and could easily be purified after extraction. In addition, the expression of foreign genes carried in the host cells could be studied and this has opened novel possibilities for biotechnological applications.

In the second part of the 1970s, new methods were developed by chemists to determine nucleotide sequences of DNA fragments^{9,10}. Hence, stepwise sequencing of longer stretches of DNA became possible thanks to the possibilities in preparing sufficient quantities of purified DNA fragments and to order such fragments by restriction mapping. The gateway to the determination of the nucleotide sequences of entire genomes was thus open in the early 1990s. In the meantime, dozens of bacterial genomes have been sequenced, ranging in their size between less than one and a few million base pairs. More recently, the DNA sequences of several eukaryotic organisms were published, including the human genome of about three billion base pairs. Compared to the early times of sequencing, this tremendous task has been highly facilitated by automatization. In the sequencing, as well as in the handling and further exploration of DNA sequences, computers play an important role and they will certainly continue to do so in bioinformatics as more and more DNA sequences will become known.

It should be mentioned here that amplification of particular DNA segments has, in the 1980s, become easier by the introduction of the polymerase chain reaction (PCR), provided that short nucleotide sequences at both ends of the DNA segment of interest are known¹¹.

The ultimate goal of biological research is to acquire knowledge on biological functions. These are often of complex nature and thus not easily accessible to experimental investigations. However, the availability of genomic DNA sequences certainly represents a first step to an improved understanding of life processes. On known DNA sequences one can search for open reading frames and for expression control signals. But to identify their real roles, it is important to apply the strategy of classical genetics to compare inherited genetic information with mutated information. For this purpose methods of site-directed mutagenesis have been developed^{12,13}. This allows one to compare in living cells, the phenotypic traits displayed by either unaltered DNA or deliberately mutated DNA. Observed differences of such traits often indicate the biological function for which the studied DNA is relevant. This general strategy of molecular genetic investigation is principally applicable to many different organisms. Once a biological function has become known, search for sequence homologies in other organisms followed by appropriate functional studies in those organisms can greatly help in amplifying our knowledge on biological functions.

Let us direct our attention for a moment to the fundamental shift of research strategies used in classical and in molecular (also called reverse) genetics. In classical genetics, studies begin with the search for mutants which are identified by an altered phenotype, hence an altered function. This variation is correlated with an alteration in the genetic information if the altered phenotype becomes transmitted to the progeny. A pairwise crossing of different variants allows the researcher to establish genetic maps which usually turn out to be linear. But so far, no knowledge on the chemical nature of genetic information is needed.

In contrast, reverse genetics starts with a segment of DNA which is sequenced. The sequence is screened for potential open reading frames and for possible expression control signals or other regulatory elements. The real functions are then explored with genetic variants prepared by site-directed mutagenesis. Therefore, reverse genetics goes from the genomic location of a particular DNA segment to its biological function, while classical genetics goes from a functional trait to its location on the genome. It is clear that the molecular genetics approach is more systematic and informative. For instance, it can open direct ways to structural studies on protein products. Note still another difference between classical and molecular genetics: In classical genetics the term 'mutation' is used for a change in a phenotypic trait, while in molecular genetics it represents an alteration in the inherited DNA sequences. It is well known that not each alteration in a DNA sequence results in an alteration of the encoded function. Therefore, special attention should be given to the definition of terms used, particularly in interdisciplinary collaborations.

As was already mentioned, the identification of particular biological functions can often lead to biotechnological applications. As more and more functions will become unravelled, the development in this field is still widely open. Of particular interest are applications in medical diagnosis and therapy, in environmental care and remediation, and in food security. In addition to these materials, often technological applications of novel knowledge, one should also pay attention to the philosophical dimension by contributing to the updating and expanding of our world-view.

The first debates on conjectural risks of genetic engineering and on their evaluations date back to the early 1970s, when the first successful experiments of *in vitro* recombination of DNA were published. The scientific community rapidly found satisfactory ways to handle such risks, and appropriate guidelines were introduced recommending special care and stepwise procedures with appropriate experimental controls. From the scientific point of view, this procedure was and still is fully in line with the expected responsibility of the research community.

As in the early times of molecular genetics, knowledge from microbial genetics on the formation of genetic variants can help us to carry out risk evaluations on a solid scientific ground¹⁴. If we compare the length of genome sequences with the linear arrangement of letters in our writing, a bacterial genome roughly corresponds to one book, while the genomes of many higher eukaryotes, including the human genome, would roughly fill 1000 books! Other genome sizes are in between these values. Genes are variable in length and may fill between a few lines to about one page. Hence, the introduction of one or a few genes into another organism is equivalent to the introduction of one sheet into the genomic library of the organism. The resulting hybrid organism is in no way novel, it will at most display some novel properties, but otherwise, it remains the kind of organism it had been before the genetic alteration¹⁵. Other types of genetic variations made by genetic engineering are even of much smaller extension. Site-directed mutagenesis usually affects only a few nucleotides. Still another genetic variation sometimes produced by genetic engineering is the reshuffling of genomic sequences, e.g. if a given open reading frame is brought under a different signal for expression control or if a gene is knocked out. All such changes have little chance to change in fundamental ways, the properties of the organism. In addition, it should be remembered that the methods of molecular genetics

themselves enable the researchers anytime to verify whether the effective genomic alterations correspond to their intentions, and to explore the phenotypic changes due to the alterations. This forms part of the experimental procedures of any research seriously carried out.

Interestingly, naturally occurring molecular evolution, i.e. the spontaneous generation of genetic variants has been seen to follow exactly the same three strategies as those used in genetic engineering¹⁴. These three strategies are: (a) small local changes in the nucleotide sequences, (b) internal reshuffling of genomic DNA segments, and (c) acquisition of usually rather small segments of DNA from another type of organism by horizontal gene transfer. However, there is a principal difference between the procedures of genetic engineering and those serving in nature for biological evolution. While the genetic engineer pre-reflects his alteration and verifies its results, nature places its genetic variations more randomly and largely independent of an identified goal. Under natural conditions, it is the pressure of natural selection which eventually determines, together with the available diversity of genetic variants, the direction taken by evolution. It is interesting to note that natural selection also plays its decisive role in genetic engineering, since indeed not all pre-reflected sequence alterations withstand the power of natural selection. Many investigators have experienced the effect of this natural force which does not allow functional disharmony in a mutated organism.

1. Avery, O. T., MacLeod, C. M. and McCarty, M., *J. Exp. Med.*, 1944, **79**, 137–158.
2. Lederberg, J., *Genetics*, 1947, **32**, 505–525.
3. Zinder, N. D. and Lederberg, J., *J. Bacteriol.*, 1952, **64**, 679–699.
4. Watson, J. D. and Crick, F. H. C., *Nature*, 1953, **171**, 737–738; 964–969.
5. Arber, W., *Annu. Rev. Microbiol.*, 1965, **19**, 365–378.
6. Arber, W., *J. Mol. Evol.*, 1991, **33**, 4–12.
7. Smith, H. O. and Wilcox, K. W., *J. Mol. Biol.*, 1970, **51**, 379–391.
8. Berg, P., *Science*, 1981, **213**, 296–303.
9. Sanger, F. and Coulson, A. R., *J. Mol. Biol.*, 1975, **94**, 441–448.
10. Maxam, A. M. and Gilbert, W., *Proc. Natl. Acad. Sci. USA*, 1977, **74**, 560–564.
11. Saiki, R. K. *et al.*, *Science*, 1988, **239**, 487–491.
12. Shortle, D., DiMaio, D. and Nathaus, D., *Annu. Rev. Genet.*, 1981, **15**, 265–294.
13. Smith, M., *ibid*, 1985, **19**, 423–462.
14. Arber, W., *FEMS Microbiol. Rev.*, 2000, **24**, 1–7.
15. Mooney, H. A. and Bernardi, G., *SCOPE 44: Introduction of Genetically Modified Organisms into the Environment*, John Wiley, Chichester, 1990.