THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

The Rise and Fall of the Idea of Genetic Information (1948-2006)

Citation for published version:
Garcia Sancho Sanchez, M 2006, 'The Rise and Fall of the Idea of Genetic Information (1948-2006)', *Genomics, Society and Policy*, vol. 2, no. 3, pp. 16-36. https://doi.org/10.1186/1746-5354-2-3-16

Digital Object Identifier (DOI):
10.1186/1746-5354-2-3-16

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Genomics, Society and Policy

Publisher Rights Statement:
Published as Open Access. © ESRC Genomics Network. www.gspjournal.com

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 27. May. 2024

GSP online Genomics, Society and Policy
2006, Vol.2, No.3, pp.16-36

The Rise and Fall of the Idea of Genetic Information (1948-2006)

MIGUEL GARCÍA-SANCHO

Abstract

On 26 June 2000, during the presentation of the Human Genome Project's first draft, Bill Clinton, then President of the United States, claimed that "today we are learning the language in which God created life".[1] Behind his remarks lay a story of more than half a century involving the understanding of DNA as information. This paper analyses that story, discussing the origins of the informational view of our genes during the early 1950s, how such a view affected the research on the genetic code (1950s and '60s) and the transformation of the information idea in the context of DNA sequencing and bioinformatics ('80s and '90s). I suggest that the concept of DNA as information reached a climax with the proposal of the Human Genome Project (HGP), but is currently facing a crisis coinciding with the questioning of the information society. Finally, I discuss the emergence of systems biology as an alternative paradigm.

Introduction

> *'The gestalt shift to information thinking in biology, with all its paradoxes and ahorias, was even more fundamental than the subsequent (1953) paradigm shift from protein to DNA…In fact, it is now hard to imagine that genes did not always transfer information or that there were other ways of knowing and doing'*[2]

The social management of genetic information is, nowadays, an extended topic of discussion. Natural and social scientists debate whether the increasingly available DNA sequences may reveal our features, should be made public to third parties or even become an object of patenting.[3] Behind all these controversies remains the idea of DNA as information, with its subsequent capacity for interpretation, storage or release, and intellectual property. Such informational understanding has penetrated so deeply the scientific and popular thinking that it is currently "hard to imagine that genes did not always transfer information or that there were other ways of knowing and doing".[4] Nevertheless, the history of DNA as information dates back to post-World War II and has significantly shaped the last 50 years of biological research.

This paper aims to uncover that story, by analysing how the informational view of DNA has influenced genetic research and clarifying what scientists have meant and mean with this concept. For that purpose, I will first review the scholarship on the origins of informational thinking in biology and its influence in the decipherment of the genetic code during the 1950s and '60s. I will then investigate the role of information in the development of DNA sequencing, how this practice made the concept shift from message to text, and how such text could be processed within the emergent field of bioinformatics in the '70s and '80s. The paper will, finally, address the origins of genomics as a discipline founded on the expectations for the scientific potential of information, and how its lack of results has resulted in increasing criticisms of the

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

concept. Such criticisms coincided with a parallel questioning of information society and the proposal of systems biology as an alternative approach to our genes.

1. Information as a message (1948-1967)

The origins of the informational conception of DNA, as well as its first applications to biology, have been extensively studied by Lily Kay. This historian has shown how after the World War II, the notion of specificity[5] – which attributed to the three-dimensional structure of biological molecules their main features - "was interchanged in a complicated manner with information", ie, the attribution of such features to an internal *code*, *text* or *message* in the molecules.[6] This shift, according to Kay, was caused by the interaction between Information Theory – a branch of engineering concerned with the transmission of signals – and genetics during the late '40s and '50s. The following sections will analyse those interactions, the concept of biological information arising from them and how it shaped, during the '50s and '60s, research on the genetic code, ie, the process by which DNA determines the structure of proteins.

The origins of informational thinking in biology

The first concept of information used by biologists, Kay argues, was borrowed from electronics and engineering.[7] Claude Shannon, Norbert Wiener and John von Neumann had, since World War II, been working on different aspects of electric information handling and their writings[8] interested a number of geneticists during the '50s. The generation and transmission of a signal could suitably model gene action, which, at that time – five years before the elucidation of the double helix – still identified with self-replicating proteins. The engineering roots of the concept resulted in a purely mathematical notion of information, not taking into account its content, but rather the accuracy of its transmission.

The dominant meaning of information in the late '40s was that derived from Information Theory, a branch of communication engineering which identified the concept with the number of options a source had when transmitting a message to a destination. Information was expressed as a quantity which depended directly on the number of available symbols and was inversely proportional to the grade of constrictions of the language. A message with a high content of information was, thus, that entailing wide possibilities in its elaboration and increased difficulties in its transmission. This led Shannon to identify the concept with the physical magnitude of entropy and to establish an estimative formula based on Boltzmann's equation. Information, according to Shannon, was entropy since a higher quantity in a message implied more uncertainty to reconstruct its meaning. Warren Weaver, co-author with Shannon of *The Mathematical Theory of Communication* – an influential book in this field – explained the concept as follows:

> *Information is a measure of one's freedom of choice when one selects a message. If one is confronted with a very elementary situation where he has to choose one of two alternative messages, then it is arbitrarily said that the information associated with this situation is unity…The two messages between which one must choose, in such a selection, can be anything one likes. One might be the text of the King James Version of the Bible, and the other might*

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

> *be "Yes". The transmitter might code these two messages so that "zero" is the signal for the first and "one" the signal for the second…Thus the two positions, closed and open, of a simple relay, might correspond to the two messages.*[9]

The content of the message was, hence, unimportant in Information Theory. This resulted in an information concept "demarcated from meaning"[10] and expressed in the form of a quantity unrelated to the material basis of the message (which, as stated above, could be the Bible or the word 'yes'). The inactivation of the meaning of information would be especially relevant when applying the concept to biology, since genetic processes are accompanied by a series of chemical reactions which were initially ignored by scientists.

One of the first researchers to apply the concept to genetics[11] was Henry Quastler. A biophysicist by background, he employed a similar equation to that of Shannon for estimating the information content of a number of biological molecules. The measurements expressed the degree of uncertainty for determining the nature and position of the atoms making up the molecules and were directly related to their structural intricacy. Proteins and enzymes, as the putative genetic material, were the first objects of his investigations, which after 1953 shifted to DNA.

Quastler was convinced that information would give an expression of the specificity of the molecules without the necessity of analysing their "reactive gadgetry", ie, investigating their chemical properties.[12] His measurements were, therefore, inspired by the mathematical and non-semantic concept of Shannon and Weaver, excluding the material – in this case chemical – content of information. Nevertheless, after numerous attempts, it emerged that the information content alone did not produce significant insights about the properties of proteins or DNA.[13] Despite this disappointment, Kay has stressed the relevance of Quastler in shaping the first investigations on the genetic code during the mid '50s.[14]

Information and the genetic code

The so-called *coding problem* arose after the elucidation of the double helical structure of DNA in 1953.[15] The confirmation of DNA as the genetic material and the attribution of a function to that substance – determining the structure of proteins – brought the notions of *code* and *message* into informational thinking. Genetic information, from then on, was viewed as the transmission of a series of instructions by DNA in the form of a coded message, which resulted in a particular sequence of amino acids in the protein. The influence of Shannon's quantitative and non-semantic concept of information, however, persisted.

All the attempts of breaking the genetic code during the mid and late '50s "blackboxed", according to Kay, the chemical basis of the process by which DNA determines the structure of proteins.[16] The researchers, following the path of Quastler and Shannon, relied on mathematical methods to find matches between nucleotide sequences and amino acid structures, rather than addressing the chemical reactions behind such correspondences. Francis Crick, one of the main investigators on the code at that time, epitomised such tendency in a 1957 paper:

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

> *This problem [coding] is a formal one. In essence, it is not concerned with either the chemical steps or the details of the stereochemistry. It is not even essential to specify whether RNA or DNA is the nucleic acid being considered. Naturally, all these points are of the greatest interest, but they are only indirectly involved in the formal problem of coding.*[17]

The formal-mathematical approach to the code led to a series of attempts to break it through cryptography between 1954 and '58. George Gamow, Francis Crick and other researchers first established the DNA nucleotide triplets as the most suitable coding input and then sought rules for reducing all the possible DNA triplets (64) to the known 20 amino acids, in order to find matches. The matches were determined by finding constants in the amino acid distribution of known protein sequences and deducing from them the coding triplets.[18] It was, thus, presupposed that no triplet coded for more than one amino acid.

Neither of these attempts led to significant results and Gamow, after laborious cryptographic studies, admitted that the "distribution of amino acids" in proteins was "purely random" and that it was, therefore, complicated to deduce patterns leading to DNA matches.[19] Despite the inconclusiveness of these investigations, Kay considers that they laid the "conceptual framework" in which the further attempts of breaking the code moved.[20] Since the late '50s, researchers have gradually opened the blackbox and analysed the chemical mechanisms of the code, though still using strong informational terminology in their papers.

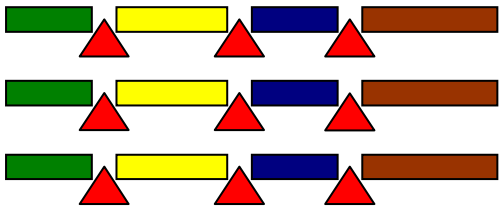
François Jacob and Jacques Monod discovered, in 1957, that the production of the enzyme βgalactosidase depends on various coordinated genes,[21] indirectly questioning the previous one-to-one triplet-amino acid assumption. The first coding match was found in 1961 by Marshall Nirenberg and Heinrich Matthaei, who in an epoch-making experiment introduced a known nucleotide sequence to a test tube and waited for a protein to be assembled. By analysing the structure of the resulting protein, they could match its amino acid sequence with the introduced nucleotides.[22] The remaining matches of the code were found within the next six years by researchers using the same strategy.

Kay and other scholars studying the genetic code see in this second stage – and especially in the persistence of informational vocabulary in the researchers' papers[23] – a transition of the information concept toward the more flexible natural languages. By addressing the chemistry of the process, scientists began manipulating the sequences of RNA, DNA and proteins, deducing from them the matches. This led them to abandon the previous mathematical meaning of information in favour of a concept closer to the idea of a text.[24] The complete transition in this direction would not, however, take place until two decades later, when information was reconfigured within DNA sequencing and bioinformatics.

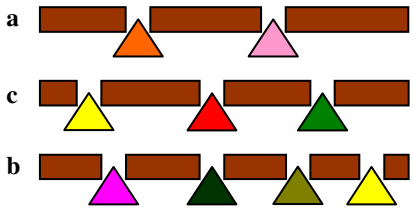
Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

Figure 1. Fred Sanger's pioneer methods to sequence proteins and DNA

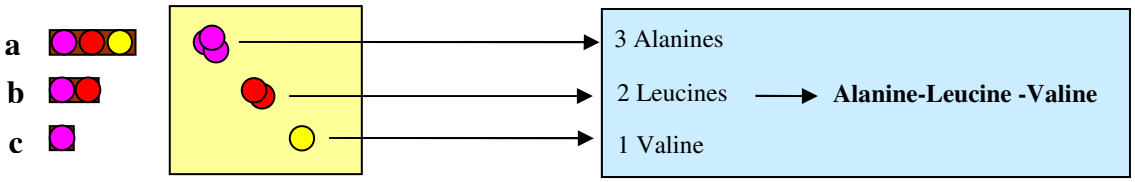
a) Degradation procedure (invented in the '40s for use on proteins)



Various samples of the protein are broken into fragments with acids (when sequencing the final amino acids) or enzymes and acids (when sequencing the whole molecule)



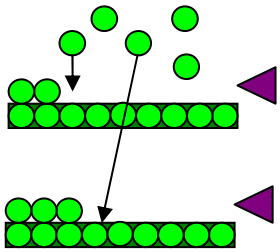
Each fragment is submitted to further breakage with in order to form overlapping sections (eg, a, c and b)



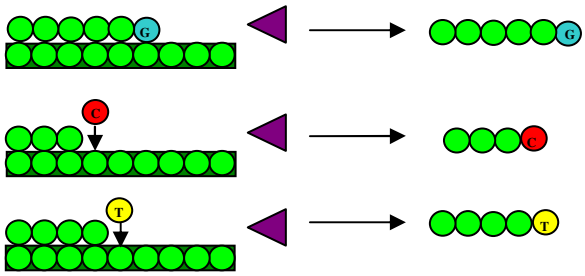
The overlapping fragments (a, b and c) are cut again and their constituent amino acids separated by paper chromatography

By identifying and quantitatively analysing the amino acids, it is possible to deduce the sequence

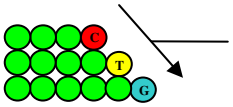
b) Copying procedure (used on DNA since 1975)



Various samples of the DNA to be sequenced were duplicated with polymerase, an enzyme which progressively incorporated nucleotides to the chains.



By different means, Sanger managed to stop polymerase selectively at adenine, cytosine, guanine and thymine.



Cytosine-Thymine-Guanine

By repeating this operation and ordering the fragments by size, the sequence could be deduced from the resulting *ladder*.

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

2. Information as a text (1962-1992)

The development of DNA sequencing marked a turning point in the concept of genetic information. The possibility of determining the linear nucleotide structure of DNA, which had been emerging since the early '70s, was based on the idea of DNA as a text. The precursors of the sequencing technologies – Fred Sanger and Walter Gilbert – presented their achievements as methods to *read* the DNA molecule. Such methods were a major factor in the transition of ongoing investigations on various organisms – *C. elegans*, yeast, *E. coli* – toward mapping and sequencing enterprises and also paved the way to the emergence of bioinformatics – ie, computer programs to process and store the growing DNA data. The following sections discuss these transformations, focusing on the new identity they created for genetic information.

Information and sequencing

Fred Sanger's research career represents a suitable case study to analyse the influence of information in sequencing, for it started before the techniques became applicable to DNA. Since 1943, Sanger developed methods to sequence proteins – particularly insulin – at the Department of Biochemistry, University of Cambridge, UK. In 1962, he moved to the recently founded Laboratory of Molecular Biology (LMB), also in Cambridge, and began applying his techniques to RNA and DNA, during the '60s and '70s, respectively.[25] Whereas in his protein research – which coincided with the first uses of the term *information* by geneticists (see above) – he was not aware of this concept, Sanger progressively assumed an informational vocabulary during his work on nucleic acids.

Sanger's methods for proteins were based on the so-called degradation procedure. This involved breaking down samples of the protein into fragments and reconstructing the amino acid sequence by matching the overlaps between the fragments of the different samples.[26] The development of his DNA techniques in the '70s required an entirely new approach to sequencing, based on copying rather than degrading the molecule. Both his *plus and minus* and *dideoxy* methods consisted of duplicating DNA with the enzyme polymerase – instead of breaking it- and successively stopping the copying process, in order to form increasingly large fragments in which the last nucleotide was known. By separating the fragments according to size – through a technique called gel electrophoresis – the sequence could be deduced from the resulting pattern (figure 1).[27]

In all his retrospective accounts, Sanger has attributed that shift of strategy to the "availability of technologies", ie, to the presence or absence of certain sequencing instruments (copying or degrading techniques, separation procedures) during his work on proteins and DNA.[28] Nevertheless, if one analyses other contemporary sequencing attempts, it is possible to see that, since the late '60s, a number of scientists were independently using the techniques Sanger employed to sequence DNA the decade after, namely fractionation gels and polymerase-based DNA copying.[29] Such 'technologies' were, therefore, available at least five years before Sanger combined them in his DNA techniques, which suggests other reasons beyond their availability for his using them.

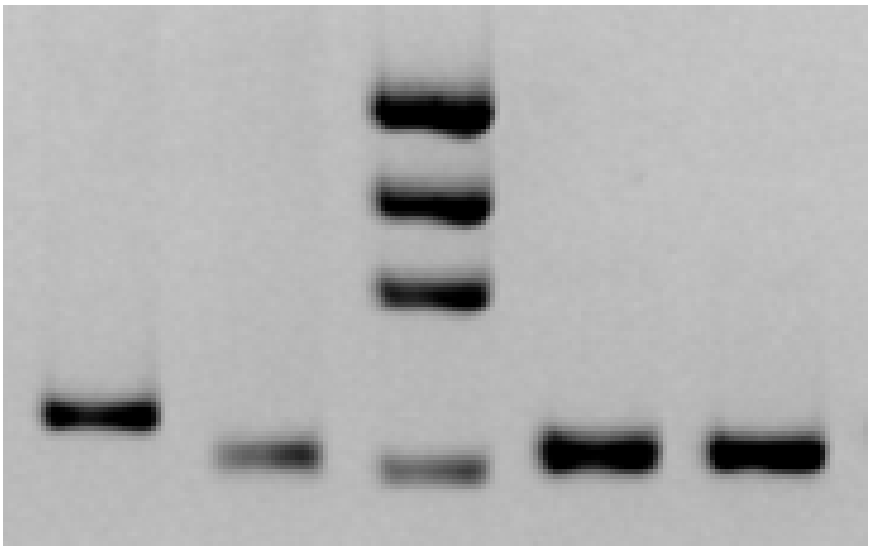
Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

The concurrence of other reasons is reinforced by significant changes of language in Sanger's papers on protein and DNA sequencing. One of the main transformations refers, precisely, to the concept of information. Whereas in the insulin papers, when he still used the degradation technique, the term only designated the sequencing results – giving information about the protein[30] – in the case of DNA, it referred to the molecule itself. An example of the latter is his Croonian Lecture, written in 1975, shortly after devising his first DNA technique:

> *DNA, the chemical component of the gene, plays a central role in biology and contains the whole information for the development of an organism, coded in the form of sequences of the four nucleotide residues. The lecture describes the development and application of some methods that can be employed to deduce sequences in these very large molecules.*[31]

This statement points to the acquisition of an informational view of DNA by Sanger, occurring gradually after his move to the LMB.[32] This acquisition could account for his transition from degradation to copying methods, especially in those instances not sufficiently explained by the availability of sequencing instruments. The perception of the genetic material as information – ie, as a text written in a biological language – may have led Sanger to articulate the already available sequencing instruments in a particular way to tackle DNA.

Figure 2. An agarose gel, similar to the ones Sanger used for sequencing. The black spots correspond to DNA fragments. (Credit: Eladio Velasco, Medical School, University of Valladolid, Spain)



Evidence supporting this idea is that in the copying procedure, the effect of the polymerase on DNA (progressively adding nucleotides to the template molecule) resembles the activity of *writing* (figure 1), while the interpretation of the gel surface by the researcher (linearly scanning the spots with the eye) mirrors the action of *reading* (figure 2).[33] In the papers describing the technique, Sanger talked about sequences which could be "read off",[34] suggesting that those analogies were more than

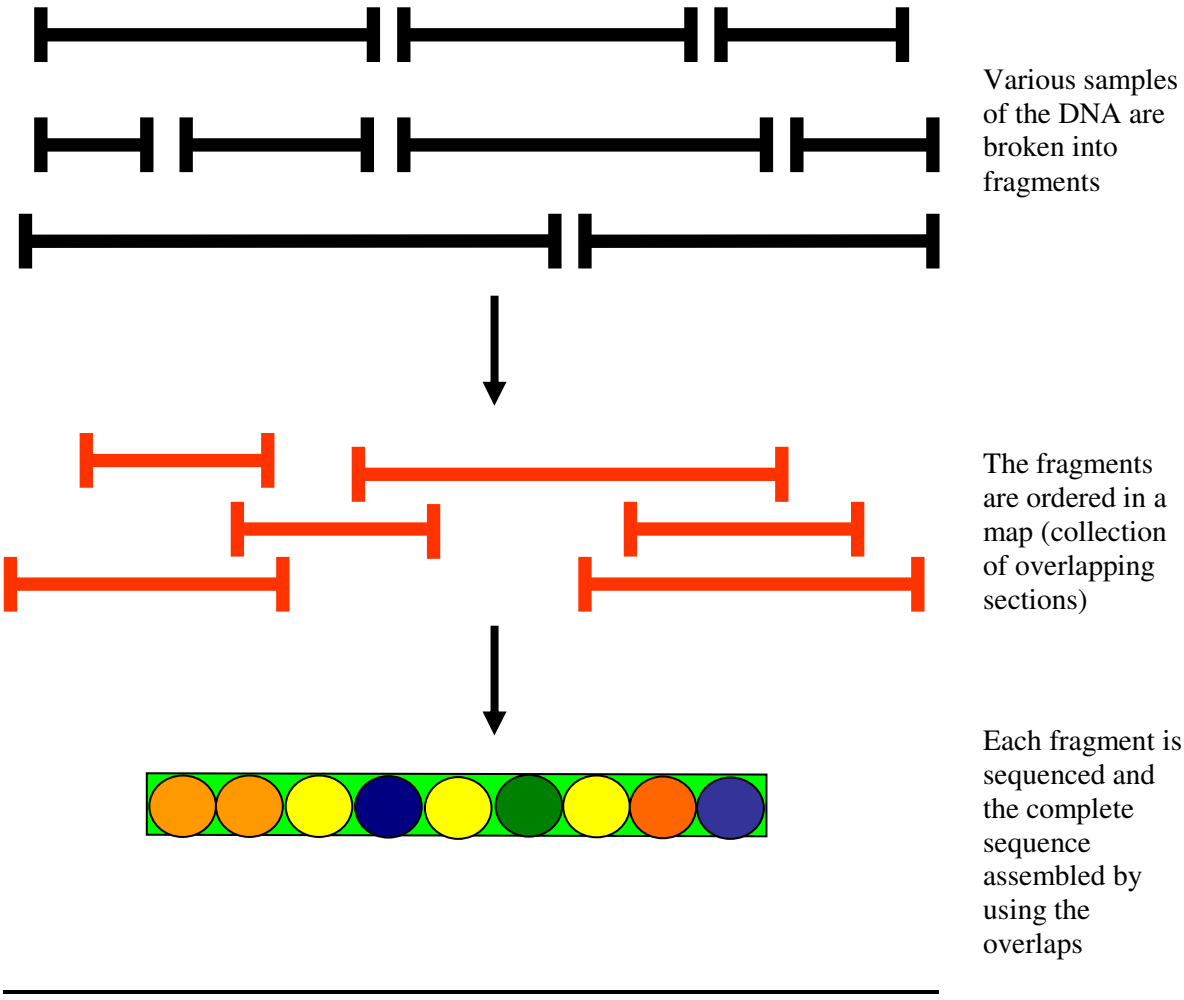
Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

accidental. It seems that the evocation of the activities of *reading* and *writing* crucially shaped the way in which Sanger employed polymerase and electrophoresis in his techniques.

Another retrospective confirmation of the importance of Sanger's informational thinking is a statement he made in 1997, twenty years after the development of his DNA techniques:

> *When I started working in the Department of Biochemistry…I thought of DNA as an inert substance. The notion of DNA containing all the information for making a complete organism would have been thought of as science fiction. But that is the way it is. The DNA is the genome and we now know how to read the information contained in it.*[35]

Figure 3. The process of DNA mapping and sequencing



Sanger's techniques had a major impact in biology, fostering the transition of multiple research projects toward sequencing. When the genomes were large, it was necessary to set a previous stage consisting in mapping the DNA. Such activity comprised

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

transforming the molecule in an ordered set of fragments which were then sequenced separately (figure 3). An example of transition towards mapping and sequencing were the investigations on the nematode worm *C. elegans*, also conducted at the LMB and initiated in the 1960s, long before Sanger's shift to DNA. The aims of the worm researchers at that time consisted of describing the nematode's development and nervous system. However, in 1983, after Sanger had developed his techniques and tested them in viruses and other simple organisms,[36] John Sulston – one of the worm scientists – decided to change the approach of the investigations towards mapping the *C. elegans* genome.[37]

Although Sulston had not explicitly admitted Sanger's influence,[38] the sequencing techniques played a crucial role in the shift of the worm research. This transition was accompanied by the move of Alan Coulson – Sanger's assistant – from the sequencing group to the *C. elegans* group at the LMB after the retirement of his boss in 1983. Coulson has considered that move a "natural step" in his career, "not completely unrelated" to his previous sequencing work.[39] In addition, the worm mapping technique introduced many of the tools Sanger had employed in sequencing, namely restriction enzymes, molecular cloning and gel electrophoresis.[40]

The first mapping results of *C. elegans* were presented in a 1986 paper, where Sulston and Coulson defined the project as "a technique for digital characterisation and comparison of DNA fragments". Genetic pieces covering 60% of the worm's genome had been analysed and reordered through the so-called "fingerprinting technique", a method similar to sequencing which also involved electrophoretic separation of overlapping DNA fragments (figure 4).[41] At the same time, researchers in the United States and Japan embarked in similar activities on yeast and *E. coli.*[42] All these investigations resulted in sequencing projects during the late '80s, with the one on *C. elegans* finishing first in 1998, two years before the presentation of the Human Genome Project's first draft.[43]

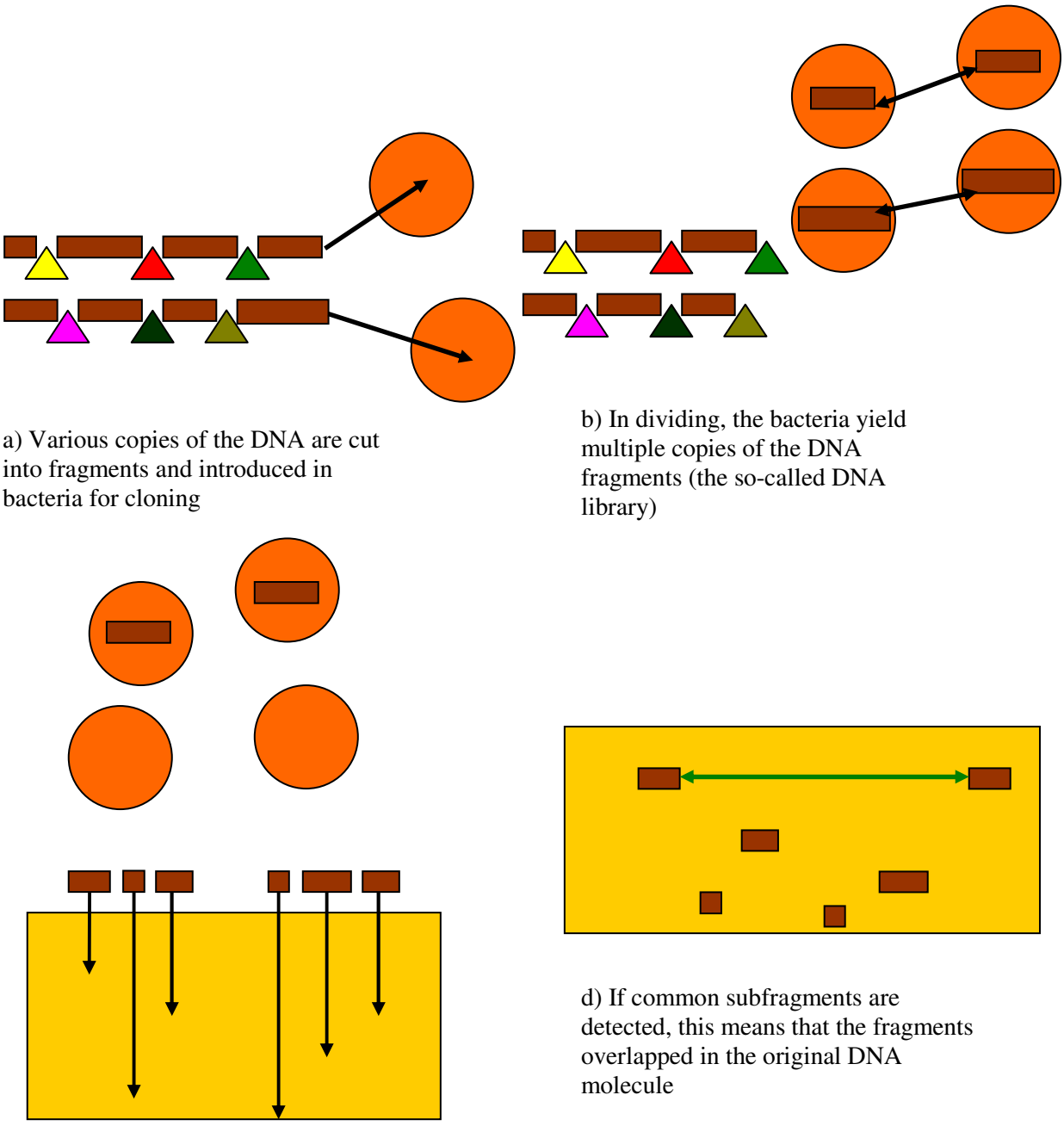
The projects on *C. elegans*, yeast and *E. coli* suggest a general shift toward mapping and sequencing in the mid '80s, induced to a great extent by Sanger's techniques. These methods did not just represent new technologies, but also a transition in the concept of genetic information from the notion of a message to that of a text. Another key factor for this transition was the emergence of computer programs to process the incoming DNA texts, also taking place during the late '70s and '80s.

DNA, text and computers

The emergence of DNA sequencing and its subsequent impact in biological research implied a transition in the concept of genetic information from that of a message with instructions to assemble proteins to that of a text written in a sequence of nucleotides. Both ideas, despite being complementary, successively focused the attention of scientists, who since the 1970s progressively shifted from the genetic code and other research on gene expression to finding means for determining and handling DNA sequences.[44] The transition from message to text influenced not only the research agendas, but also the techniques devised by scientists. In this context, sequencing and its associated software were examples of the new text-like conception of DNA.

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

Figure 4. The 'fingerprinting' mapping technique



a) Various copies of the DNA are cut into fragments and introduced in bacteria for cloning

b) In dividing, the bacteria yield multiple copies of the DNA fragments (the so-called DNA library)

c) The fragments are broken again and the resulting subfragments separated trough electrophoresis (a technique in which the smaller move faster and the bigger slower)

d) If common subfragments are detected, this means that the fragments overlapped in the original DNA molecule

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

Understanding DNA as a message implied a concern with its function rather than with its structure. This was especially suitable for the research on the genetic code, in which the aim of the scientists was to determine how the genetic material acts – ie, how it transmits instructions to assemble proteins. The techniques devised by the investigators at that time (1950s and '60s) were based on cryptography or biochemistry, and directed to find the matches between DNA and proteins rather than their sequences (see *Information and the genetic code* above). Goals and techniques were different when DNA began being conceived as a text during the '70s. At that time, the object of researchers turned to DNA structure and the methods were oriented to obtain or process its sequence, not directly addressing genetic function.[45] Sanger's techniques and its associated computer programs emerged as representatives of this view and progressively became dominant throughout the '80s.

The first sequencing software developed in a context of rapid achievements in computer science. One of the main areas of expansion was dynamic programming, a strategy consisting of decomposing complex problems into subroutines (eg, a, b, c, d and e), many of which overlapped in the sense that they solved their task (a) and many others (b and d). That way, it was possible to design software leading the computer to overcome the problem with the minimum number of subroutines (a, c and e).[46]

Dynamic programming showed remarkable similarities with mapping and sequencing large genomes. The latter techniques also implied breaking a long unit (various samples of DNA) into fragments, ordering the fragments in overlapping sections and sequencing each fragment, using the overlaps to assemble the whole sequence (figure 3). These similarities led a group of researchers to embark in program design first for protein and then for DNA sequencing between the '70s and early '80s.[47]

One of the pioneers in the DNA software was Rodger Staden, a mathematical physicist also based in the LMB and attracted to the sequencing group in the late '70s, after the development of the DNA techniques.[48] Staden's early programs were designed for Sanger's virus-based first sequencing projects and described DNA as a series of "character strings". They were devoted to "ordering" and "storing…gel reading data", and permitted sequences to be edited, searched and compared. After processing, the sequences were saved in magnetic disks as files which were measured in kilobytes.[49]

In 1982, Staden launched an improved line of software for sequence assembling.[50] Its main feature was a command for rapidly comparing sequences which had been already proposed by the French researchers Jean-Pierre Dumas and Jacques Ninio. The command involved estimating all the possible combinations between seven nucleotides (16,348) and placing them in a matrix in the form of single digits (combination-1, combination-2, etc, up to combination-16,348). The sequences to match were then scanned by the computer only once, determining which combination numbers they had. If a digit coincided in both sequences, there was, consequently, an overlap.

This strategy, according to Dumas and Ninio, derived from a series of protocols used in "classical lexicographic search problems" as "hash coding and separate chaining" – ie, the common word-processing operation of searching for terms beginning with the same letters: *pupp** for *puppet* and *puppy*.[51] The way in which Staden described it in his 1982 paper clearly shows that he was conceiving and processing DNA as a text:

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

> *The sequence can be considered as a series of 7 character words that overlap each of their neighbours by 6 characters. Using an alphabet of 4 [the number of DNA nucleotides] it is possible to make 16,348 (4 to the power of 7) different words of length 7. If we have a table of 16,348 positions we can record whether or not each of the possible words of length 7 are present in the consensus [assembled sequence]…In order to compare a gel reading with the consensus we simply look in the table to see if any of the gel's 7 character words are present.*[52]

The DNA software had a crucial role in the development of the mapping and sequencing projects during the '80s. Sulston and Coulson soon adjusted Staden's programs to perform mapping as well as sequencing, and adopted them for the *C. elegans* genome map. Since 1984, the worm's DNA fragments were processed as a series of position "data", which were stored in a central "data base".[53] This considerably eased the international cooperation arising around the project, since all the laboratories involved in *C. elegans* shared the database information.[54] The production and exchange of information became, hence, the main feature of the nematode research.

The widespread use of computers in mapping and sequencing, together with the expansion of these activities during the '80s, shows how the reduction of DNA to information was increasingly perceived as a significant scientific goal at that time. The first debates about the feasibility of the HGP, beginning in 1985,[55] confirmed such perception. Nevertheless, as the human and other DNA sequences grew, an increasing scepticism about their utility arose among researchers and scholars. This led to an on-going debate on the adequacy of the information concept in biology.

3. Climax, blows and alternatives to information (1990-present)

The period between the proposal of the HGP and its launch (1985-90) was marked by the enthusiasm of scientists and society, both strongly believing in the potential of genetic information. This stage was followed by a growing scepticism toward the concept, as the increasingly available human DNA sequences revealed insufficient information to bring insights about how genes worked.[56] It is interesting to see that this scepticism coincided with critiques, among sociologists, of the information society as a distinctive social structure characterising the '80s and '90s. The following sections will discuss such coincidence and the alternatives to genetic information currently emerging in the natural and social sciences.

Genomics and the HGP

The emergence of the HGP and parallel projects (eg, *C. elegans*) fostered the configuration of a new discipline – genomics – engaged with the mapping and sequencing of large DNA molecules. The term was coined by the geneticist T.H. Roderick in 1986, after a meeting with his colleagues Victor McKusick and Frank Ruddle,[57] who, the following year, launched *Genomics,* a journal devoted to endeavours in this field.

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

The first issue of *Genomics* defined the discipline as resulting from "a marriage of molecular and cell biology with classical genetics", being also "fostered by computational science". McKusick and Ruddle remarked in the journal's editorial on the importance of various types of "map" as objects of "biological significance".[58] Such maps allowed the gathering of information about development and disease, the main object of the new discipline:

> *Mapping all expressed genes…regardless of whether their function is known, sequencing these genes together with their introns, and sequencing out from these is seen by many as 'the way to go'. The ultimate map, the sequence, is seen as a rosetta stone from which the complexities of gene expression in development can be translated and the genetic mechanisms of disease interpreted. For the newly developing discipline of mapping/sequencing (including analysis of the information) we have adopted the term GENOMICS.*[59]

The emergence of genomics marked a period of optimism and trust in the power of genetic information. The benefits for health were the most widely perceived, with researchers talking about a "new era of molecular medicine". In 1992, two years after the launching of the HGP, Leroy Hood – one of its main proponents – predicted that medicine would move "from a reactive mode (curing patients already sick) to a preventive mode (keeping people well)".[60] Genetic diagnosis and gene therapy were seen as the means for that transformation, the former permitting the determination of sequence mutations causing the diseases – even before the birth of the patient – and the latter ensuring the substitution of faulty genes for healthy ones.[61]

Not all the analyses of the HGP were, however, positive at that time. Also in 1992, philosophers of biology Sahotra Sarkar and Alfred Tauber published a paper challenging the excessive euphoria toward human DNA sequencing.[62] According to them, the HGP proponents' expectations were "naive", since there was no "practicable way to begin characterising biological function at the molecular level from sequence information alone". The paper showed how DNA sequences were insufficient to deduce key biological events as protein folding. It also questioned the idea of a *human genome*, since there are important differences between the DNA sequences of every particular person.[63]

Tauber and Sarkar's critique entailed a subsequent challenge to the idea of genetic information. If the DNA sequences, as they argue, were limited deductive tools, the aim of computing biological function from a molecular text[64] became impossible, or at least more difficult to achieve than expected. This, according to them, made the HGP the culmination of the reductionist tradition which had shaped biological research in the last two centuries:

> *The HGP is the ultimate product of an extreme reductionist vision of biology that has held that to understand better one needs only to go smaller. Reductionism in molecular biology constitutes a research program that attempts to explain and understand biological systems in terms of the physical interactions of their parts. From that point of view, it is natural to assume that fundamental understanding of biology comes only from the level of DNA, the alleged blueprint for living systems.*[65]

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

The critical perspectives on the HGP spread as the project progressed throughout the '90s. The difficulties researchers encountered during their efforts to deduce features from the increasingly available DNA sequences reduced the perceived potential of such sequences to explain biological function. This has made the promised transformations in health only partially fulfilled today, six years after the publication of the human sequence.[66] Whereas in diagnosis the envisioned transition towards preventive medicine is becoming a reality – with the predisposition to cancer and other conditions being increasingly detected from mutations[67] – the situation is more problematic in gene therapy. The obstacles to introducing the replacement genes into cells and to the use of viruses as vectors (gene-carriers) has resulted in the failure of many clinical trials, some of them with patients suffering from collateral infections.[68]

Such difficulties, however, have not prevented some scholars defending the persistence of genetic information. Biologist and historian of science Michel Morange has claimed that the criticism of genes as containing insufficient information "to explain its organisation and functioning" derives from the consideration of only "one level" of biological organisation. Given the complexity of life, integrated by a "hierarchical structure" of macromolecules, organelles, cells, organs and organisms, the analysis of any biological phenomenon should address the "relation between the complexity of a hierarchical system and that of its components". Within such complex organisation, Morange claims, the gene remains "indispensable for modern biology".[69]

Morange's remarks lead to the perspective of systems biology, emerging in the life sciences especially after the presentation of the HGP. The main postulation of this new approach is situating DNA not at the centre of biological function, but within a complex network of interactions with other cell and environmental components. It is interesting that systems biology has developed in a context of simultaneous critique of information society, the belief among sociologists that information was the main means of social knowledge and power. The rise-and-fall dynamics of information society since the '80s shows remarkable parallels with those of genetic information and the subsequent emergence of alternative paradigms.

Information society and systems biology

The rise of genomics and the first proposals of the HGP were accompanied by important developments in informatics. The personal computer and the workstations, invented in the '70s, expanded during the next decade among a population that increasingly wished to become individual users.[70] This gradually led to a view of information processing as an essential driving force in society, a trend quickly recognised by scholars who set about analysing the social role of information.[71]

Nevertheless, the key investigation in this regard was conducted by the sociologist Manuel Castells. In a three-volume treatise, he argued that, since the late '70s, the development of information technologies[72] had been altering the material basis of society and creating a "new way of producing, communicating, managing and living". This had resulted in the information society, a structure in which the "generation, processing and transmission of information" became "the fundamental sources of productivity and power". One of the key features of this social form was "the

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

networking logic of its basic structure", since markets and population groups tended to be electronically interconnected.[73]

The period described by Castells – comprising the 1980s and early '90s – coincided with the golden age of genetic information, in the light of genomics and the launching of the HGP. Nevertheless, a number of critical responses followed the publication of his book in 1996. Frank Webster and Brian Winston accused Castells of technological determinism, by stressing that society had also played a role in accommodating the new informational instruments.[74] David Edgerton, in a more recent investigation, considers the information society a "fashionable, and misleading", term since our civilisation is characterised by the massive production of material entities. Examples of this are "the sheer bulk" of objects in homes, the "unprecedented weight of stuff in the shops, the piles of paper in any office, not to mention the proliferation of computers, fax machines, and Xeroxes".[75] Edgerton's critique suggests that there is a material counterpart to information which should be taken into account when analysing its social role.

The challenges to the information society have coincided with new explanatory models of gene action in the natural sciences. In the same year that the HGP was presented (2000), Hood founded the Institute of Systems Biology,[76] a centre proposing a new approach to DNA and, more generally, to cell function. Systems biology considers DNA a unit in permanent contact with external environmental inputs and other biological molecules in the process of gene expression. This makes it necessary to take into account all the components in the system and to define their interactions, in order to find emergent properties not deducible from the isolated elements.[77]

The way in which Hood describes systems biology makes it compatible with an informational understanding of DNA. This researcher talks about "digital information" encoded in the genetic material and interacting with "environmental information" of a rather analogical nature.[78] Other accounts of systems biology have distinguished different "levels of information" – gene, protein, cell, tissue, viscera and the body – postulating "interfaces" that convert information of one level into the others.[79] This new approach to biology remains, hence, highly informational, but the fact that it distinguishes between different sorts and levels of information makes it different from genomics.

Such differences are reflected in a recent paper where David Eisenberg and other researchers still consider biology as concerned with the "study of transmission of information" by "informational macromolecules – DNAs, RNAs and proteins". The aim of these investigators is to "compute" and "simulate" cellular function, as well as the "three dimensional structures" of proteins from those informational inputs. Nevertheless, the authors admit that "other types of genomic and proteomic information", apart from DNA, will be necessary for such a task. They conclude by postulating a "genome-wide information" repository, which "will continue to expand" with further research.

Eisenberg and his colleagues' perspective reflects the changes to the idea of information triggered by systems biology. First, this new approach includes types of information derived not only from DNA, but also from proteins and other biological

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

molecules. Second, genetic information – ie, the DNA sequence – is no longer the centre of biological function, but a component at the same level of and in permanent interaction with enzymes, nutritional inputs and the environment. This mode of operation, in which biological function cannot be *read* from the sequence structure of one of the components, does not mirror a text any more, but rather the network structure of computers interconnected through the Internet. The coincidence of systems biology with the critiques of information society suggests that a new understanding of our social and biological structures is now emerging.

Conclusions

This paper has shown a historical transition in the concept of genetic information from the idea of a message to the idea of a text between the 1970s and '90s. This transition resulted in a shift of focus in the understanding of DNA from that of a molecule containing the instructions to assemble proteins to that of a precise sequence of nucleotides. It was also central to the evolution of biological research from the genetic code and other forms of genetic expression (1950s and '60s) to DNA sequencing and bioinformatics (1970s to '90s). The conception of DNA as a text also shaped the emergence of genomics in the late '80s as a research culture engaged with the reduction of this molecule to map and sequence information.

The idea of DNA as alternatively a message and a text mirrors the two concepts of the gene proposed by the philosopher of biology Lenny Moss. This scholar distinguishes between a "preformationist gene (Gene-P)" which would "predict phenotypes but only on an instrumental basis" and an epigenetic gene (Gene-D) which would act as a "developmental resource that provides possible *templates* for RNA and protein synthesis".[80] Moss's concept of Gene-P as a gene that is essentially functional and concerned with phenotypic effects would correspond with the idea of DNA as a message – a molecule containing instructions shaping the features of the organism. His concept of Gene-D, meanwhile, would connect with DNA as a text, ie, a sequence informing other biological molecules.

Moss demonstrates a "conflation" of both concepts of gene during the 20th century, caused by the emergence of the "rhetoric" of DNA "as code and information". Such rhetoric led to genes being conceived "at once" as both "molecular sequences and pieces of the phenotype" shaping the organism's features. This, according to Moss, resulted in an essential misunderstanding of the concept of gene. Despite genes being understandable as both structural templates and forces influencing the organism, these functions are performed by different biological mechanisms. It is, therefore, incorrect to consider DNA a text-like "blueprint which regulates its own execution" or functioning.[81]

The conflation demonstrated by Moss complements my own story. The transition I have discussed between the genetic code, sequencing and bioinformatics (from message to text, from Gene-P to Gene-D) was not accompanied by a complete shift on the conceptual side. Researchers did not completely abandon the previous notion of DNA as a message when they shifted to text-like sequencing and genome processing in the '70s and '80s. This led DNA to be seen as both a sequence containing information (text, Gene-D) and as information shaping the phenotype (message, Gene-P). The high

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

expectations of genomics and the HGP the following decade were, then, founded on this double perspective, ie, the possibility of deducing phenotypic effects from the sequence.

Systems biology qualifies such beliefs and helps to situate DNA within the confines of the Gene-D concept. The fact of DNA being studied in its interaction with other molecules maintains its template dimension, but limits considerably its operative power. The coincidence of this transition with the critiques of information society suggests, additionally, that raw information is not currently seen by people as a means of knowledge and power. In both biology and society, information needs to be interconnected and, especially, interpreted to avoid *saturation.*

The scientific and social power of information has, therefore, been qualified in the advent of the 21st century. The control of voluminous amounts of data no longer guarantees a privileged position, since the data should be previously processed to produce knowledge and power. In biology, this processing is achieved by connecting DNA with such non-genetic inputs as proteins or the environment.[82] This opens a new horizon in research – the so-called post-genomic era – but at the cost, perhaps, of an old dream of scientists: that of reading our destiny in our genes.

Acknowledgements

John Lagnado (UK Biochemical Society Archives); Frances Martin and Joan Green (Sanger Institute Library); Annette Faux and Michael Fuller (Laboratory of Molecular Biology of Cambridge Archives and Storage); Wellcome Trust Archives and Manuscript Collection; UK National Archives; Andrew Mendelsohn, David Edgerton, Andrew Warwick and Robert Iliffe (Centre for the History of Science, Imperial College, London); Soraya de Chadarevian (University of California Los Angeles); Fred Sanger, John Sulston, Sydney Brenner, Bart Barrell, Richard Durbin, Alan Coulson (Sanger Institute and Laboratory of Molecular Biology of Cambridge); Adam Bostanci and other researchers at the ESRC Centre for Genomics in Society (University of Exeter, UK); José Manuel Sánchez Ron, Javier Ordóñez, Antonio Sillero, Rafael Garesse and other researchers (Universidad Autónoma de Madrid); Eladio Velasco (University of Valladolid, Spain).

This research was conducted while holding postgraduate fellowships given by Caja Madrid Foundation, Madrid City Hall and Residencia de Estudiantes (Spain) as well as a Hans Rausing Fellowship given by the Centre for the History of Science, Imperial College, London. Without them, it would not have been feasible.

---

[1] B. Clinton. Remarks on the completion of the First Survey of the entire Human Genome Project 2000: 2. Available at http://clinton5.nara.gov/WH/New/html/genome-20000626.html [Last accessed 6 December 2006]

[2] L. Kay. 2000. Who Wrote the Book of Life? A history of the genetic code. Stanford University Press: xv.

[3] See D. Kevles. Out of Eugenics: the historical politics of the human genome. D. Nelkin. The social power of genetic information. E. Lander. DNA fingerprinting: science, law and the ultimate identifier. H.T. Greely. Health, insurance, employment, discrimination and the genetics revolution. All in D.J. Kevles and L. Hood (eds.) 1992. The Code of Codes: Scientific and Social Issues in the Human Genome

© ESRC Genomics Network. www.gspjournal.com

Project. Cambridge. Harvard University Press. See also E. Fox Keller. Is there an organism in this text? In P.R. Sloan (ed.) 2000. Controlling our destinies. Indiana. University of Notre Dame Press. See also C. Hamilton. Biopiracy and genes as information. A. Bostanci. The genomics revolution and patenting. Both papers at the Postgraduate Forum on Genetics and Society 9th Annual Colloquium. University of Cardiff, 2005.

[4] Kay, op. cit. note 2, p.xv.

[5] One of the first proponents of the idea of specificity was the biochemist Emil Fischer, who in the late 19th century formulated the lock-and-key hypothesis, according to which molecules interacted by adjusting their three-dimensional shapes. This theory was developed by Linus Pauling in the late '30s and '40s, attributing to the three-dimensional structure of antibody proteins their precise recognition of pathogens. L. Pauling. A theory of the structure and process of formation of antibodies. Journal of the American Chemical Society 1940; 62: 2643-2657.

[6] Kay, op. cit. note 2, p.5. For further research on the influence of the information concept in biology see S. Sarkar. Biological information: a sceptical look at some central dogmas in molecular biology. S. Sarkar, ed. 1996. The Philosophy and History of Molecular Biology: New Perspectives. London. Kluwer Academic: 187-233. E. Fox Keller. 1995. Refiguring life: Changing Metaphors in 20th Century Biology. New York. Columbia University Press. M. García-Sancho. 2004. The Configuration of Genetic Information and Language during the Second Half of the 20th Century. Constructing and Representing and Object through Genetics. MSc dissertation available at the Centre for the History of Science, Imperial College, London. L. Moss. 2003. What Genes Can't Do. Cambridge. MIT Press. J. Maynard Smith. The concept of information in biology. Philosophy of Science 2000; 67: 177-194. P. Griffiths. Genetic information: a metaphor in search of a theory. Philosophy of Science, 68. C. Brandt. Genetic code, text, and scripture: metaphors and narration in German molecular biology. Science in Context 2005; 18(4): 629-648.

[7] Kay, op. cit. note 2, chapter 3.

[8] C. Shannon and W. Weaver. 1948. The Mathematical Theory of Communication. Chicago. University of Illinois Press. N. Wiener. 1948. Cybernetics: or Control and Communication in the Animal and the Machine. Cambridge. MIT Press. J. von Neumann. The general and logical theory of automata. In A.H. Taub, ed. 1948. Collected Works, vol. 5: 288-328. Erwin Schrödinger had previously introduced the idea of a "code-script" associated to genetic activity. E. Schrödinger. 1944. What is Life? Cambridge University Press edition: 61-62.

[9] Shannon and Weaver, op. cit. note 8,p.100.

[10] Kay, op. cit. note 2, p.95. García-Sancho, op. cit. note 6,pp.12-13.

[11] Another area of biology in which the mathematical concept of information was applied (more successfully than in genetics) was brain science. Eg, A. Rosenblueth, N. Wiener, J. Bigelow. Behaviour, purpose and teleology. Philosophy of Science 1943; 10: 18–24.

[12] S. Dancoff and H. Quastler. The information content and error rate of living things. H. Quastler, ed. 1953. Essays on the Use of Information Theory in Biology. Chicago. University of Illinois Press: 263-273.

[13] Quastler organised two symposia on the use of Information Theory in biology in 1952 and '56. In the latter, he explicitly admitted that the application of this concept "to the study of living things" had not produced "many results so far". This inconclusiveness persisted in 1963, after Quastler's premature death. H. Quastler. The status of Information Theory in biology. H.P. Yockey, ed. 1958. Symposium on Information Theory in Biology. London. Pergamon Press: 399.

[14] Kay, op. cit. note 2, p.123. For further discussion see García-Sancho, op. cit. note 6, pp.18-19.

[15] Watson and Crick's second paper on the double helix stated that "the precise sequence of bases" in DNA was the "code" which carried "the genetic information". J. Watson and F. Crick. Genetical implications of the structure of deoxyribose nucleic acid. Nature 1953; 171: 964

[16] See Kay, op. cit. note 2, chapter 4.

[17] F. Crick, J.S. Griffith and L.E. Orgel. Codes without commas. Proceedings of the National Academy of Sciences 1957; 43: 416-21.

[18] Crick et al, op. cit. note 17. G. Gamow and M. Ycas. The cryptographic approach to the problem of protein synthesis. H.P. Yockey, ed. 1958, op. cit. note 13, pp.63-69.

[19] G. Gamow, A. Rich and M. Ycas. The problem of information transfer from nucleic acids to proteins. Advances in Biological and Medical Physics 1956; 4: 60.

[20] L. Kay. A book of life? How a genetic code became a language in P.R. Sloan, ed., op. cit. note 3, pp.115-124.

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

[21] A.B. Pardee, F. Jacob and J. Monod. The genetic control and cytoplasmic expression of inducibility in the synthesis of galactosidade by *Escherichia Coli*. Journal of Molecular Biology 1959; 1: 165-178. F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology 1961; 3: 318-56.
[22] M. Nirenberg. and H. Matthaei. The dependence of cell-free protein synthesis in *E. Coli* upon naturally occurring or synthetic polyribonucleotides. Proceedings of the National Academy of Sciences 1961; 47: 1588-1602.
[23] Jacob and Monod referred to the process of protein synthesis as "cytoplasmic expression", whereas Nirenberg and Matthaei defined as "meaningful information" the nucleotide sequences introduced in the test tube. Pardee, Jacob and Monod, op. cit. note 23, p.165. Nirenberg and Matthaei, op. cit. note 24, p.1601.
[24] Kay, op. cit. note 2, chapters 5 & 6. Sarkar, op. cit. note 6. Fox Keller, op. cit. note 6. García-Sancho, op. cit. note 6, p.27 onwards.
[25] For an autobiographical account of Sanger see F. Sanger. Sequences, sequences and sequences. Annual Review of Biochemistry 1988; 57: 1-28.
[26] F. Sanger. Chemistry of insulin. Science 1959; 129: 1340-1344.
[27] F. Sanger et al. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology 1975; 94: 441-8. Sanger et al. DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Sciences 1977; 74: 5463-7.
[28] F. Sanger. Interview with author. Sanger Institute. Cambridgeshire. 2005. Sanger. op. cit. note 27, pp.20-24.
[29] R. Wu and A. Kaiser. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. Journal of Molecular Biology 1968; 35: 523-37. A.C. Peacock and C.W. Dingman. Molecular weight estimation and separation of ribonucleic acids by electrophoresis in agarose-acrylamide composite gels. Biochemistry 1968; 7(2): 668-74.
[30] F. Sanger. The terminal peptides of insulin. Biochemical Journal 1949; 45: 563–574.
[31] F. Sanger. The Croonian Lecture. Proceedings of the Royal Society of London Series B 1975; 191: 317-33.
[32] Sanger's interaction with molecular biologists as Crick or Sydney Brenner was crucial for his acquisition of an informational conception of DNA. Crick and Brenner were two of the main proponents of genetic information and attracted Sanger to the LMB for using sequencing in the context of the genetic code. This has led Soraya de Chadarevian to interpret Sanger's shift from proteins to RNA and DNA as a transition from biochemistry (representative of his work on insulin) to molecular biology (representative of his work at the LMB). S. de Chadarevian. Sequences, conformation, information: biochemists and molecular biologists in the 1950s. Journal of the History of Biology 1996; 29: 361-86.
[33] M. García-Sancho. Technologies for reading. The influence of the information concept in Fred Sanger's sequencing methods for insulin and nucleic acids (1943-1977). Postgraduate Forum on Genetics and Society 9th Annual Colloquium. University of Cardiff. 2005.
[34] Sanger, 1975, op. cit. note 29, p.443. Sanger 1977, op. cit. note 29, p.5463.
[35] F. Sanger, quoted in L. Fletcher and R. Porter. A Quest for the Code of Life: Genome Analysis at the Wellcome Trust Genome Campus. London. Wellcome Trust: 72.
[36] Sanger applied his techniques to the bacteriophage viruses ØX-174 and λ – both with small genomes - during the late '70s and early '80s.
[37] The nematode project and its transition towards mapping and sequencing has been object of extended scholarship. See S. de Chadarevian. Of worms and programmes: *C. elegans* and the study of development. Studies in the History and the Philosophy of the Biological and Biomedical Sciences 1998; 29(1): 81-105. S. de Chadarevian. Mapping the worm's genome: tools, networks, patronage. In H.J. Rheinberger and J.P. Gaudillière, eds. 2004. From Molecular Genetics to Genomics: The Mapping Cultures of Twentieth Century Biology. London. Routledge. See also R. Ankeny. The natural history of *C. Elegans* research. Nature Review Genetics 2001; 2: 6-10.
[38] Sulston has rather attributed his decision of mapping *C. elegans* to himself acquiring, in the mid '80s, a global perspective of the worm's genome, conceiving it as a whole unit susceptible to be fully characterised. J. Sulston. 2002. The Common Thread: A Story of Science, Politics, Ethics and the Humane Genome. London. Bantam: 38-39.
[39] A. Coulson. Interview with author. Laboratory of Molecular Biology. Cambridge. 2005

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

[40] M. García-Sancho. Technologies for reading: the influence of the information concept in the research on the worm *C. elegans*. British Society for the History of Science Annual Conference. University of Kent. Canterbury. 2006.
[41] A. Coulson, J. Sulston et al. Toward a physical map of the genome of the nematode *C. Elegans*. Proceedings of the National Academy of Sciences 1986; 83: 7821-25.
[42] M. Olson et al. Random-clone strategy for genomic restriction mapping in yeast. Proceedings of the National Academy of Sciences 1986; 83: 7826-30. Y. Kohara. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell 1987; 50: 495-508.
[43] *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans:* a platform for investigating biology. Science 1998; 282: 2012-8.
[44] García-Sancho, op. cit. note 33, p.7 and onwards. García-Sancho. op. cit. note 40, pp.17-21.
[45] The fact of the sequences not directly leading to DNA function would be crucial for the further development of biological research (see below).
[46] For more information about dynamic programming see T. H. Cormen et al. 2001. Introduction to algorithms. Cambridge. MIT Press. Chapters 15 and 32.
[47] S. Needleman and D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 1970; 48: 443-53. D. Sankoff. Matching sequences under deletion / insertion constraints. Proceedings of the National Academy of Sciences 1972; 61: 4-6. T.F. Smith and M.S. Waterman, M.S. Identification of common molecular subsequences. Journal of Molecular Biology 1981; 147: 195-97.
[48] Staden had previously worked in the Structural Studies Division of the LMB, which employed computer programs since the '50s to perform calculations in x-ray crystallography. For more details about his move to the sequencing group see F. Sanger and M. Dowding, eds. 1996. Selected Papers of Frederick Sanger (with Commentaries). London. World Scientific: 344. See also M. Levitt. The birth of computational structural biology. Nature Structural Biology 2001; 8: 392 - 393 (especially: 392-93).
[49] R. Staden. Sequence data handling by computer. Nucleic Acids Research 1977; 4(11): 4038. R. Staden. A strategy of DNA sequencing employing computer programs. Nucleic Acids Research 1979; 6(7): 2601. R. Staden. A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Research 1980; 8(16): 3673.
[50] R. Staden. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Research 1982; 10(15): 4731–4751.
[51] J.P. Dumas and J. Ninio. Efficient algorithms for folding and comparing nucleic acid sequences. Nucleic Acids Research 1982, 10(1): 197.
[52] Staden, op. cit. note 50, p.4743.
[53] Coulson, Sulston et al, op. cit. note 41, p.7821.
[54] García-Sancho, op.cit. note 40, pp.20-21.
[55] Such debates started with a meeting in Santa Cruz attended by Sulston. Brenner was a UK representative in the commission of the US National Academy of Sciences subsequently analysing the HGP. R. Cook-Deegan. 1994. The Gene Wars: Science, Politics and the Human Genome. London. W.W. Norton and Company. Chapter 5.
[56] The enthusiasm and further scepticism toward the HGP mirrored the reception of the idea of the genetic code among researchers between the '60s and '70s. Its decipherment in 1967 triggered the interest of biologists and even linguists in the concept of genetic information and DNA acting as a message (F. Jacob, R. Jakobson, C. Lévi-Strauss and P. L'Héritier. Vivre et parler. Les Lettres Françaises 1968. R. Jakobson. Linguistics in its relation to other sciences. A. Graur, ed. 1969. Actes du Xe congrès international des linguists. Bucharest. Editions of the Academy of the Socialist Republic of Rumania: 75-111). However, the decade after, those same investigators expressed their reservations on the utility of the linguistic models to account for genetics (F. Jacob. Le modele linguistique en biologie. Critique, 322: 197-205). Erwin Chargaff was one of the only scientists critical of the idea of the code since the '60s. E. Chargaff. 1963. Essays on nucleic acids. Amsterdam. Elsevier. Chapters 7 and 8.
[57] B. Kuska and B. Beer. Bethesda and biology: how genomics came into being. Journal of the National Cancer Institute 1988; 90: 93.
[58] V. McKusick and F. Ruddle. Editorial: a new discipline, a new name, a new journal. Genomics 1987; 1: 1.
[59] McKusick and Ruddle, op. cit. note 58, p.1.
[60] C.T. Caskey. DNA-based medicine: prevention and therapy & L. Hood. Biology and medicine in the twenty first century. Both in D.J. Kelves and L. Hood, eds. 1992, op. cit. note 3.

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com

[61] The belief in the medical potential of the HGP remained almost a decade after, during the presentation of the project's draft (2000). There, an exultant Clinton claimed that the human DNA sequence would "revolutionise the diagnosis, prevention and treatment of most, if not all, human diseases", including Alzheimer's, Parkinson's, diabetes and cancer. Such pathologies would benefit from "new drugs" targeting "the faulty genes". Clinton, op. cit. note 1, p.2.
[62] Four years after his challenge to the HGP, Sarkar wrote a critique of the idea of information linked to the genetic code (Sarkar, op. cit. note 6).
[63] A. Tauber and S. Sarkar. The Human Genome Project: has blind reductionism gone too far? Perspectives in Biology and Medicine 1992; 35(2): 220-35.
[64] One of the most enthusiastic proponents of the computational approach to biology - and of the aim of many of the criticisms of Tauber and Sarkar - was the co-inventor of DNA sequencing Walter Gilbert. W. Gilbert. A vision of the grail. In Kelves and Hood, eds., op. cit. note 3.
[65] Tauber and Sarkar, 1992, op. cit. note 63, p.228. The debate on the utility of the HPG and, more generally, the concept of genetic information persists nowadays. One of its more recent manifestations has been a paper by the evolutionary biologist John Maynard Smith, which is supportive of the information idea, and a critical response by the philosopher of biology Paul Griffiths. Maynard Smith and Griffiths, op. cit. note 6.
[66] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860-921. C. Venter et al. The sequence of the human genome. Science 2001; 291: 1304-51.
[67] See, for instance, E. Esteban-Cardeñosa; M. Duran; M. Infante.; E. Velasco; C. Miner. High throughput detection method to scan BRCA1 and BRCA2 based on heteroduplex analysis by capillary array electrophoresis. Clinical Chemistry 2005, 50(2): 313-320.
[68] C.E. Thomas et al. Progress and problems with the use of viral vectors for gene therapy. Nature Reviews Genetics 2003; 4: 346-58. H. Parekh-Olmedo. Gene therapy progress and prospects: targeted gene repair. Gene Therapy 2005; 12: 639-46.
[69] M. Morange. 2001. The Misunderstood Gene. Cambridge. Harvard University Press: 161-62 and 2-6.
[70] P. Ceruzzi. 2000. A History of Modern Computing. Cambridge. MIT. Chapters 8-9.
[71] See, for instance, E.B. Parker. Implications of new information technology. Public Opinion Quarterly 1974; 37(4): 590-600. A. Danzin. 1979. Science and the Second Renaissance in Europe. Oxford. Commission of the European Communities.
[72] Castells explicitly refers to genetic engineering as a form of information technology, in the line of the above discussed informational thinking of biologists. M. Castells. 1996. The Information Age. Vol. 1: Economy, Society and Culture. Oxford. Blackwell: 47-50.
[73] Ibid, vol. 1, pp.5 and 21.
[74] F. Webster. Is this the Information Age? Towards a critique of Manuel Castells. City 1997; 8: 71-84. B. Winston. 1998. Media Technology and Society: A History from the Telegraph to the Internet. London. Routledge.
[75] D. Edgerton. 2006. The Shock of the Old. Technology and Global History since 1900. Oxford University Press: 96-97.
[76] See www.systemsbiology.org/
[77] L. Hood. The networks of life, Schrödinger Lecture. Imperial College, London. 2004. L. Hood et al. In press. Biological Information and the Emergence of Systems Biology. Denver. Roberts & Co.
[78] Hood, op. cit. note 77. Other current biological research line which remains highly informational is biosemiotics, represented by Jesper Hoffmeyer and Claus Emmeche. These scholars defend DNA nucleotides being signs, ie, informational entities in the same sense as letters and sounds in written and oral languages. C. Emmeche. A semiotical reflection on biology, living signs and artificial life. Biology and Philosophy 1991; 6: 325-340. J. Hoffmeyer. 1996. Signs of Meaning in the Universe. Indiana University Press.
[79] R. Kitney. A revolution in medicine: the role of information technology in the future of medicine. Graduate School Guest Lecture. Imperial College, London. 23rd November 2006. Morange, op. cit. note 69, pp.2-3 and 159-163.
[80] Moss, op. cit. note 6, p.xiv.
[81] Moss, op. cit. note 6, pp.44-50 and chapter 2.
[82] New biological findings as the existence of overlapping genes, introns or distant interacting genome regions are, furthermore, showing that DNA does not exactly act as a text. This has led scholars at the University of Pittsburgh to assess the different definitions of the gene concept, seeking a new-contemporary one (see http://www.pitt.edu/~kstotz/genes/genes.html).

Genomics, Society and Policy, Vol.2 No.3 (2006) ISSN: 1746-5354
© ESRC Genomics Network. www.gspjournal.com