From the Genetic to the Computer Program

THIS IS AN ADVANCED DRAFT OF A PUBLISHED PAPER.

REFERENCES AND QUOTATIONS SHOULD ALWAYS BE MADE TO THE PUBLISHED VERION, WHICH CAN BE FOUND AT:

García-Sancho M. (2012) "From the genetic to the computer program: the historicity of 'data' and 'computation' in the investigations on the nematode worm *C. elegans* (1963-1998)", *Studies in History and Philosophy of Biological and Biomedical Sciences, 43:* 16-28.
URL: http://dx.doi.org/10.1016/j.shpsc.2011.10.002

From the genetic to the computer program: the historicity of 'data' and 'computation' in the investigations on the nematode worm *C. elegans* (1963-1998)

Miguel García-Sancho

Department of Science, Technology and Society

Spanish National Research Council (CSIC)

Calle Albasanz, 26-28

28037 – Madrid

SPAIN

Current email: *miguel.gsancho@ed.ac.uk*

Abstract:

This paper argues that the history of the computer, of the practice of computation and of the notions of 'data' and 'programme' are essential for a critical account of the emergence and implications of data-driven research. In order to show this, I focus on the transition that the investigations on the worm *C. elegans* experienced in the Laboratory of Molecular Biology of Cambridge (UK). Throughout the 1980s, this research programme evolved from a study of the genetic basis of the worm's development and behaviour to a DNA mapping and sequencing initiative. By examining the changing computing technologies which were used at the Laboratory, I demonstrate that by the time of this transition researchers shifted from modelling the worm's genetic programme on a mainframe apparatus to writing minicomputer programs aimed at providing map and sequence data which was then circulated to other groups working on the genetics of *C. elegans*. The shift in the worm research should thus not be simply explained in the application of computers which transformed the project from hypothesis-driven to a data-intensive endeavour. The key factor was rather a historically specific technology – in-house and easy programmable minicomputers – which redefined the way of achieving the project's long-standing goal, leading the genetic programme to co-evolve with the practices of data production and distribution.

Keywords: *C. elegans*, genetics, computer, program, software, data, genomics, model organism.

-1.Introduction

Historians and philosophers of biology have amply investigated the use of *C. elegans* as a model organism for biomedicine. This tiny nematode worm, first proposed in the early 1960s, has become one of the preferred objects for STS scholarship on model organisms, together with mice, the *Drosophila* fly and the weed *Arabidopsis thaliana* (Löwy and Gaudillière, 1998; Rader, 2004; Kohler, 1994; Leonelli, 2007ab; see also Burian, 1993; Geison and Laubichler, 2001). The literature on *C. elegans* addresses the selection of this worm over rival species and its subsequent impact on the development of molecular biology (Ankeny, 2001; de Chadarevian, 2000, 2002, pp. 287 and ff.). It also explores the origins of an initiative to map the worm's genome in the mid 1980s and how, the decade after, *C. elegans* became a pilot project for the sequencing of the human genome (de Chadarevian, 1998, 2004). The emergence of *C. elegans* databases as shared information resources for the community of investigators working on the worm has also been a main line of inquiry (Leonelli and Ankeny, this volume; Leonelli, 2010b).

A pressing philosophical question has been the use of *C. elegans* as a "descriptive model" for biomedical researchers. Rachel Ankeny has documented how the proposal of the worm as a model organism occurred at a "preexplanatory phase" in which its role was not to test theories or hypotheses, but to serve as an ideal type to provide detailed descriptions. This descriptive usage developed during the 1970s and shaped the early research on *C. elegans*, during which exhaustive data about its development and nervous system was gathered. The worm's extended descriptions, according to Ankeny, were "essential prerequisites" to pose "explanatory questions" about its development and behaviour. These explanatory questions,

however, were not necessarily implied during the early descriptive modelling of *C. elegans* (Ankeny, 1997, 2006, quotes from 2000, pp. S62 and S67).

The accumulation of descriptive information on the worm and other model organisms has fostered the construction of centralised data banks. Ankeny and Sabina Leonelli have investigated the emergence of community databases around different model organisms, among them *C. elegans*, whose WormBase became a referent for those working on the nematode in the 1990s (Leonelli and Ankeny, this volume). Due to these databases being shared by a whole community of investigators and their increasing size, the standardisation and easy exchange of data must be strictly observed. This makes *C. elegans* and other community databases suitable objects for the development of bio-ontologies interlinking different collections of data and easing their circulation from one context to another (Leonelli, 2010a).

WormBase is the materialisation of a long trajectory of research on *C. elegans* which has been a main concern for historian Soraya de Chadarevian. She identifies as a major shift in the worm project the abandonment of the hypothesis of a "genetic programme" which allegedly governed *C. elegans* development and behaviour. This hypothesis was postulated between 1973 and 74 by Sydney Brenner, a researcher at the Laboratory of Molecular Biology of Cambridge (LMB, UK) and first proponent of the worm as a model organism. It presupposed a direct and unmediated connection between *C. elegans* genes, and their behavioural and developmental effects. During the first half of the 1980s, the genetic programme became increasingly questioned and this led to a renegotiation of the investigations on the worm. As a result of this, the *C. elegans* mapping and then sequencing

project emerged and the worm became the first multicellular organism with a fully sequenced genome (de Chadarevian, 1998, 2004).

A common point of all the scholarship on *C. elegans* is the acknowledgment of the role that computational models have played in the investigations on the worm. The *C. elegans* nervous system, as Ankeny has shown, was conceptualised as a "wiring diagram" in which the connections between the neurons were visualised by comparison with an electric circuit (Ankeny, 2001, pp. S262 and ff.). De Chadarevian argues that the hypothesis of a "genetic programme" and Brenner's early experiments on the worm were modelled on the way computers ran software instructions (de Chadarevian, 1998, p. 88). The WormBase, for Leonelli and Ankeny, represents both the meeting point and a powerful shaping factor of past and present collaborations between the *C. elegans* community (Leonelli and Ankeny, this volume).

This universal acknowledgement contrasts with the comparatively meagre investigation of the computing technologies that *C. elegans* researchers used at each stage of the project. De Chadarevian and Ankeny have explored the specialised software that Brenner's team designed to deduce the neuron connectivity of the worm and later to construct a physical map of its genome (de Chadarevian, 1998, 2004; Ankeny, 2001). However, little is known about the computers the team had at hand, the way researchers used them, and alternative apparatus and usages at other divisions of the LMB. De Chadarevian, building on previous historiography, refers to an "information discourse" which originated in cybernetics and, from the late 1940s onwards, shaped genetic research, including that on *C. elegans* (de

Chadarevian, 1998, p. 89). The particular technologies and practices in which such information discourse materialised are still largely unknown.[1]

This paper explores the computing technologies and practices that the *C. elegans* researchers adopted from the start of the experiments on the worm, in the mid 1970s, to the completion of the sequence of its genome (1998). The incorporation of mainframe computers into the scientific activity of the LMB has been reconstructed by de Chadarevian, but she does not explicitly link it to the *C. elegans* project (id., 2002, ch.4). I will make such connection and extend my previous research on the introduction of minicomputers into the LMB, with particular emphasis on the adaptation of DNA sequencing software to the worm's mapping effort (García-Sancho, forthcoming, ch.3). My argument will be that the changing apparatus and computing applications which operated at the LMB between the 1960s and 90s had a transforming effect on the *C. elegans* project, being a decisive factor in its transition towards mapping and sequencing. They shaped the way in which the worm researchers understood the practice of computation, the notion of programme and how the descriptive data about *C. elegans* should be collected, stored, distributed and used.

The impact of computing technologies on *C. elegans* research suggests that during the late 1980s and 90s, this project shifted from hypothesis-driven to a data-intensive endeavour. The community of worm researchers multiplied and its organisation and work became increasingly dependent on data gathering technologies. As in other genomic initiatives emerging at the same time – among them the Human Genome Project – the researchers' efforts were founded on the belief of the power of information to achieve fundamental

scientific insights. The completion of these initiatives and the subsequent proliferation of map and sequence information have resulted in the current perception that biology has become a data-driven science.

However, if we look back at the history of *C. elegans* as a model organism, information and computing technologies have always played a key role in the research efforts. The description of the worm's development and behaviour, conducted long before the mapping and sequencing initiatives, involved the accumulation of large amounts of data, a substantial part of which was processed with the help of computers. The project was soon developed by an internationally disperse community and its founding hypothesis – that the worm was governed by a genetic programme – was inspired in the functioning of computers. It is thus not clear what was new with the computer-based collection, storage and distribution of *C. elegans* map and sequence data.

An answer to this question requires to historically unpack the notions of 'computer', 'programme' and 'data'. By following the computing technologies used by the worm researchers and the data they produced, I will show to what extent the aims and hypotheses behind the *C. elegans* project were transformed, in line with the shift of biomedicine towards genomics and data-intensive endeavours. This historical reconstruction will complement other perspectives presented in this special issue and contribute towards its overall aim: to critically appraise the proposal by biomedical researchers and computing companies of a supposedly new data-driven science. I will also build on other recent attempts of establishing a historical perspective on genomics and, particularly, on the

compilation of data around *C. elegans* and other model organisms (Suárez-Díaz, 2010; Ankeny, 2010; Müller-Wille and Charmantier, this volume).

-2.The worm as a computational tool

Previous scholarship on *C. elegans* has shown how in 1963, Brenner proposed the use of a close variant of this nematode – *C. briggsae* – to address the genetic basis of development and behaviour. The proposal was part of an application to extend and fund future lines of research at the LMB one year after its foundation (de Chadarevian, 1998; Ankeny, 1997, ch.2; see also Brenner, 1963ab). It was formulated at a time in which molecular biologists were looking for new research problems and moving towards eukaryotic organisms after determining the mechanisms of gene action in bacteria and viruses (Morange, 1997; Yi, 2008).

Brenner's initiative, in this regard, was not an original one and he acknowledged in his proposal that "many other molecular biologists (…) were thinking along similar lines" at that time. The really important issue of his proposed project was not so much to investigate a higher organism, but to define clearly "the nature of the problem" to address and "the right experimental approach". However, both the research problem and the experimental approach were conditioned by the higher organism selected by Brenner. Nematode worms were his final choice, due to them allowing the reduction of "complicated phenomena" to "simple units", much as molecular biology had successfully done in the past (quoted in de Chadarevian, 1998, p. 82).[2]

Brenner's experimental approach was specified in the mid 1970s, when he had already shifted from *C. briggsae* to *C. elegans*. Between 1973 and 74, he published two manifesto papers which stated that *C. elegans* behaviour was governed by a genetic "programme". With programme, Brenner meant a set of "simple" rules which connected the worm's genes with certain visible behavioural traits. Behaviour was, thus, defined as the "result of a complex set of computations" initiated in the genes and "performed by the nervous system", according to the specifications of the genetic programme. Both the programme and the performed computations possessed a "logical structure" which could be determined experimentally. All the cellular and environmental elements external to the genes but mediating behaviour were considered a "separate question" at that moment (Brenner, 1974a, 1973, quotes from pp. 296-271).

This approach shaped Brenner's first experiments on *C. elegans*, initiated in 1974 and inspired in his previous research on the genetic code, conducted in viruses and bacteria during the 1950s and 60s. The *C. elegans* experiments started with the isolation of a large number of behavioural mutant worms and the association of such mutations to different groups of genes. Brenner then conducted complex crossings between the nematodes and determined which mutations were inherited together. This permitted him to construct a "map" with six "linkage groups" or series of associated genes corresponding with the worm's six chromosomes. By measuring the frequency of joint inheritance, Brenner also determined which mutation genes were closer to each other in the chromosomes (higher frequency) and which farther (lower) (Brenner, 1974b, quotes from pp. 84-91).

INSERT FIGURE 1

Brenner's approach, according to Ankeny, allowed him to establish *C. elegans* as an "idealised" or "abstract entity" which represented the genetics of behaviour. His early experiments were followed by detailed observations in which Brenner and other members of his group traced the divisions and movement of cells during the worm's development and the different degrees of connectivity between its neurons. The aim was to associate the genes of the linkage map with data from the observations – i.e. a mutated gene with alterations in the cell division pattern of development or neuron circuitry. Ankeny argues that the establishment of *C. elegans* as a "descriptive model" derives from this experimental design (Ankeny, 2000, p. S67; 1997, chs.3-4).[3]

The association of mutated genes with developmental and behavioural alterations was guided by Brenner's hypothesis of a genetic programme. This hypothesis, for de Chadarevian, was "imported into molecular genetics from electronic computing as part of a more general information discourse". Building on the work of other historians (see footnote 1), de Chadarevian shows how notions such as programme, information, code or message "had become currency" in the LMB and Brenner had acquired an "own computer" for his research activity by the end of the 1960s. In one of his manifesto papers, Brenner distinguished between the "hardware" of *C. elegans* and its "software", stating that the project would concentrate in the former, i.e. the general connections between genes and behaviour, without entering into the detailed biochemical reactions which led to a particular behavioural mutation (de Chadarevian, 1998, pp. 88-89).

Brenner's wording, together with the long-lasting influence of cybernetics in genetics (Kay, 2000; Sarkar, 1996; Fox Keller, 1995) make de Chadarevian's argument plausible. However, a question left open by her account is the technological materialisation of the information discourse on which Brenner modelled the characteristics of his genetic programme. This is a historically relevant question, since given the rapid development that informatics were experiencing at that time, Brenner's notion of programme would have changed depending on the concrete source of his modelling. A tentative answer is the extended tradition of computer usage at another LMB division investigated by de Chadarevian: that devoted to X-ray crystallography.

During the early 1950s, before the foundation of the LMB and when most of its members were based in the Cavendish Laboratory, X-ray crystallographer John Kendrew established a cooperation with the Mathematical Laboratory of the University of Cambridge. This laboratory had designed EDSAC, a mainframe computer to assist other researchers at the University. The computer had enormous dimensions (see Figure 2b) and was operated with punched cards submitted to the Mathematical Laboratory by the user researchers (de Chadarevian, 2002, ch.4). In 1952, Kendrew and Cambridge mathematician John Bennett designed a suit of programs to perform the arithmetical operations "of addition and subtraction" in one and a half milliseconds, and "multiplication in six milliseconds" (Bennett and Kendrew, 1952, p. 109).

Those programs were adopted by Kendrew and other Cambridge crystallographers to solve in a manageable time the tens of thousands of calculations required to deduce the three-dimensional structure of a molecule after its X-ray analysis. Researchers submitted to the

EDSAC operators punched cards with the outcomes of the analyses and the mainframe operators ran the programs to perform the necessary operations, mainly Fourier syntheses and Patterson projections. The programs solved the calculations in hours or days, depending on the volume of data. The results were then returned to the Cavendish and allowed any trained crystallographer to model the position of the atoms in the molecule, according to its electron density. Kendrew and his colleague Max Perutz extensively used mainframe calculations to determine the three-dimensional structure of myoglobin and haemoglobin, an achievement for which they were awarded the Nobel Prize in 1962 (de Chadarevian, 2002, pp. 111-132; see also García-Sancho, forthcoming, ch.3).

The use of the computer persisted after the move of the Cavendish crystallographers to the LMB and became a sign of identity of the Laboratory's X-ray crystallography division. Despite shortly after the move (1962) in-house minicomputers began being introduced into the LMB, the Cambridge crystallographers remained using mainframes "for large calculations". Up to the late 1970s, they regularly shipped punched cards and paper tapes to the EDSAC-2, located miles away from the LMB, and an IBM at Imperial College, London (de Chadarevian, 2002, quote from p. 130).

Brenner had joined the Cavendish Laboratory in 1957 and, as Kendrew, was further transferred to the LMB. The similarities between the genetic programme he postulated and Kendrew and Bennett's programs for crystallographic calculations suggest that Brenner's computer modelling of *C. elegans* derived from the long-standing tradition of mainframe use by the Cavendish crystallographers. The functioning of both programmes was based on the transformation of an input according to a series of logical rules – arithmetic rules in the

case of crystallographers and genetic mechanisms in Brenner's research. As a result of these transformations, an output which satisfied the objectives of the investigators was produced – the electron densities of the molecules or the behavioural mutations in the case of Brenner.

This one-to-one input-output relationship epitomised the operation of mainframe computers, whose hardware characteristics led them to respond in a similar fashion independently of the software they run; the software only specified the logical rules according to which the input would be transformed into a particular output. The hardware-software differentiation was in Brenner's second manifesto paper, where he specified how far he intended to go in the determination of the genetic programme: he would only investigate the "hardware" relationship between mutations and behavioural effects without addressing the "software" biochemical transformations which mediated between the worm genes and phenotype (Brenner, 1974a, p. 787).

The conceptualisation of *C. elegans* as a mainframe survived the introduction of in-house computers into Brenner's group. During the early 1970s, shortly after acquiring his own computer, Brenner employed PhD student John White, who had previously worked as an electrical engineer. Brenner and White sought to incorporate computing technologies to the worm project at a time in which biomedical laboratories were proving increasingly permeable to the emerging minicomputers (November, 2006, chs.4-5; 2004). In this regard, White's thesis was devoted to the design of a "computer system" for the automation of the description of *C. elegans* neuron connectivity. It consisted in a graphics terminal which allowed the visualisation of images and a series of programs which were run in Modular

One, a pioneer minicomputer commercialised by the British company Computer Technology Limited and adapted for use by small groups of operators in laboratories or offices (White, 1974, p. 6).

White's design enabled the digitisation of the electron microscope images on which Brenner's group based its observations of the worm's nervous system. The images represented groups of neurons and suggested connections between them in different manifestations of *C. elegans* behaviour. The minicomputer programs attempted to reconstruct processes from comparison of multiple consecutive pictures which were previously displayed in the graphics terminal (ibid., pp. 27 and ff.). These graphics devices were beginning to be used by crystallographers to model the three-dimensional structure of molecules from mainframe data (Francoeur and Segal, 2004; Wieber, 2006). However, in White's use the Modular One showed important memory and speed limitations to process large amounts of images. Despite the problem being partially solved with the alteration of the operating system, most of the worm's wiring diagram needed to be determined by hand (de Chadarevian, 1998, p. 90).

The data generated by White's computer system was intended to match neuron connectivity patterns with their genetic specification. The system, thus, represented a particular computer use within the project rather than an attempt to model *C. elegans* behaviour. This shows that White and Brenner's use of minicomputers did not question the conceptualisation of the worm as a mainframe, neither the organisation of the project around the hypothesis of a genetic programme. The Modular One under White's program was rather directed to provide evidence in support of this central hypothesis. It was with the

gradual questioning of the genetic programme in the 1980s when minicomputers at Brenner's group started to play a remarkably different role.

-3.Sequencing software and the transition towards DNA mapping

The investigations of eukaryotic organisms in which Brenner and other molecular biologists embarked during the 1970s showed their first experimental results at the beginning of the following decade. They reflected increasing exceptions to the mechanism of gene action embodied in the genetic programme. This led to a growing questioning of the hypothesis throughout the 1980s, even in the *C. elegans* group (de Chadarevian, 1998, pp. 94 and ff.). The results of the description of the worm's development and nervous system, with which Brenner sought to link mutated genes with observed phenotypic effects, reflected innumerable qualifications to the expected invariant rules.

John Sulston, the member of Brenner's group in charge of the description of development, concluded in 1983 that the mechanisms which governed cell division in the worm possessed "numerous exceptions" and this provided cells with a certain degree of "autonomy" (Sulston et al, 1983, quote from p. 110). Similarly, Brenner and White stated three years later that "a knowledge of the detailed structure" of a nematode's nervous system did not in itself "provide any answers" to the genetic origins of neurons or their connections (White et al, 1986, p. 58). In light of these results, Brenner became increasingly doubtful of the genetic programme and publicly admitted that the principles of

organisation in higher organisms could not be “embodied in a simple chemical device”, as it was the case “for the genetic code” (Lewin, 1984, p. 1327).

The criticisms to Brenner’s hypothesis are considered by de Chadarevian to be crucial in the “renegotiation” that *C. elegans* research experienced in the mid 1980s. The dismissal of a simple relationship between genes, development and behaviour led Brenner’s group to change the orientation of the project and embark in a different description of the worm. This transformation was shaped by two circumstances at the LMB: firstly, “new techniques, in particular recombinant DNA technologies and DNA sequencing” were at that time spreading in the Laboratory (de Chadarevian, 1998, p. 97). Secondly, Brenner left the LMB in 1986 and new researchers were incorporated to the *C. elegans* group.

The most significant incorporation, for de Chadarevian, was that of Alan Coulson. Coulson had worked in the neighbouring LMB division of Frederick Sanger in the development of techniques to determine the nucleotide sequence of DNA of different organisms (id., 2004, pp.101 and ff.; see also García-Sancho, 2010). These techniques had been released during the mid and late 1970s, and in the subsequent years Sanger and Coulson tested them in the genome of different viruses (Sanger and Coulson, 1975; Sanger, Nicklen and Coulson, 1977; García-Sancho, forthcoming, ch.3). Following Sanger’s retirement in 1982, it was decided that Coulson would move to the worm group and start a cooperation with Sulston. By that time, Sulston had decided to leave the description of *C. elegans* development and initiate the construction of a physical map of the worm’s genome.

Sulston and Coulson, with the help of Brenner and another *C. elegans* group member, Jon Karn, adapted Sanger's sequencing techniques and invented the "fingerprinting method". Due to the size of the worm's genome – of approximately 100 million nucleotide pairs – Sanger's sequencing techniques could not be directly applied to such a large DNA molecule. The fingerprinting method, instead of determining the worm's nucleotide sequence, allowed the reconstruction of its genome from multiple overlapping DNA fragments (Coulson et al, 1986; see also Sulston and Ferry, 2002, pp. 57 and ff.).

INSERT FIGURE 2a

Another historically relevant incorporation not highlighted by de Chadarevian was that of Rodger Staden. Staden had a background in mathematical physics and started his career as a technician in the LMB crystallography division. His duty was to design computer programs to process the X-ray data from which structural calculations were conducted. During the late 1970s, he was approached by members of Sanger's group, who proposed him a cooperation for the computer handling of DNA sequences. Staden subsequently joined Sanger's laboratory and built an academic career as the LMB researcher in charge of the computer programs for DNA sequencing.

A main novelty of those programs was that they were designed for in-house computers instead of mainframes. Staden first used the PDP-11 minicomputer, which was commercialised by the US company Digital Equipment Corporation (DEC) shortly after the launch of Modular One. In 1981, the PDP-11 was replaced by the next generation of DEC computers, the VAX 11/780. Both apparatus consisted in a keyboard attached to a

processor which executed the instructions of the user. They were purchased with funds from the Medical Research Council (MRC, the body of the British Government which administered the LMB) and fitted in a specific room which was shared by various LMB teams.[4] Minicomputers, according to Staden, would allow the sequencing researchers to take direct and immediate responsibility for the editing of their data, rather than submitting punched cards to remote mainframes and awaiting their return (Moody, 2004, p. 15; García-Sancho, forthcoming, ch.3).

INSERT FIGURE 2b

Staden's academic career in Sanger's group started with the pursuit of a PhD on *Computer Methods to Aid in the Determination and Analysis of Nucleic Acid Sequences.* In his thesis, submitted in 1984, he compared the DNA nucleotides – adenine, cytosine, guanine and thymine – with an "alphabet" of four characters. He further defined nucleotide sequences as "series of 7 character words". The main objective of Staden's programs was to find overlapping regions in the partial DNA fragments derived from a sequencing project. Those fragments were divided into continuous seven nucleotide sets and, when overlaps were detected, assembled into a "consensus sequence" representative of the virus or other organism to which the project was directed (Staden, 1984a, quote from p. 73). This strategy led Staden to import algorithms from the then emerging word processing software, especially those which permitted to conduct interactive searches within a text (García-Sancho, forthoming, ch.3). Staden's use of the computer was, thus, markedly different from that of the LMB crystallographers: rather than calculating molecular structures, he sought to find nucleotide patterns within DNA sequences.[5]

Also in 1984, Staden published a new program to partially automate the entrance of sequence data into minicomputers. The program, called GELIN, was based on a "digitiser" formed by a special pen and two microphones. This device was manufactured by the company Science Accessories and commercially named Graphbar. It was used to register drawings, graphs and other types of images into computers and visualise them via graphics terminals.

In GELIN, the image to digitise was the autoradiograph, a film with a black and white pattern of bands which was the outcome of Sanger's sequencing techniques (see Figure 2b). The user touched each band with the pen and the device, by recording the sound with the microphones, determined the bands' position on the film. GELIN translated this position information into a particular DNA nucleotide and thus progressively gathered the sequence (Staden, 1984b, pp. 502-503). It was used in combination with the other programs described in Staden's thesis, consequently allowing the members of Sanger's group to determine the DNA fragments contained in autoradiographs and assemble them into a consensus sequence.

Following the launch of the *C. elegans* mapping initiative, Staden started to cooperate with Sulston and Coulson, whom he knew from Sanger's group. The fingerprinting method was also based on autoradiographs and this enabled Staden to adapt the sequencing programs to assemble overlapping DNA fragments into a physical map rather than a consensus sequence. The programs, which were adapted in cooperation with Sulston and other LMB researchers with expertise in computing, became an integral part of the worm project. In

1985, Brenner requested to the MRC the acquisition of a VAX 8600 – faster and with more memory than the 11/780 – which was subsequently used in mapping *C. elegans* and in other LMB projects.[6] Staden moved to the worm group and became a full member of the team.

In 1986, Sulston, Coulson, Brenner and Karn presented the initial results of the *C. elegans* mapping project. They defined the fingerprinting method as a technique "for digital characterisation and comparison of DNA fragments", and described a "data base" in which information about the fragments and their overlaps was being stored. As in Staden's sequencing programs, the authors referred to the overlaps between the fragments as "contigs" (Coulson et al, 1986, quotes from pp. 7821 and 7823). The aim of the project, as that of the sequencing programs, was to reduce the worm's genome to a single contig, i.e. an uninterrupted set of overlapping DNA fragments.

Both the database and the mapping "software" were further described in two papers published in 1988 and 89 in *Computer Applications in Biosciences*, a journal which was first released in 1985. The autoradiographs were digitised either via Graphbar or a "custom-built scanning densitometer" which was manufactured at the LMB workshop (Sulston et al, 1989, p. 101). This latter device gradually became the standard in the mapping technique, enabling to automatically scan the film and register the bands into the minicomputer without manipulation of a pen. The bands were then visualised in a graphics terminal and their position determined by comparison with a pre-recorded autoradiograph, which acted as a control.[7]

A “package” of programs allowed the user to locate overlaps in the visualised bands and define the contigs. The program MAPSUB determined the matches between the bands and CONTIG9 was used to estimate the degree of overlap and accept or reject the contig. If accepted, the contig was incorporated into the database and visualised on the screen (see Figure 2b). A cursor enabled to further edit the map, refining the position of the contig, deleting it or adding text tags. Sulston, Staden and the other members of the worm team considered that the “judgement of the operator”, who had access to the “primary data”, was “more powerful” than a fully automated approach (Sulston et al, 1988, quotes form pp. 125 and 132).[8]

The operation of the mapping software contrasted with the previous use of the computer in the *C. elegans* project. Whereas in the 1970s Brenner had modelled the worm in a mainframe apparatus to compute its genetic programme, Sulston, Coulson and Staden were now writing minicomputer programs to gather mapping data about the worm’s genome. Staden’s incorporation to the worm group was essential for this shift. Unlike Brenner and White, he was unconnected with the hypothesis of a genetic programme and saw the computer as a working instrument rather than a model. This led him to introduce minicomputers for their better convenience in the assemblage of DNA sequences and then *C. elegans* mapping data.

This use of minicomputers transformed the way in which Sulston and Coulson conceptualised the term ‘programme’. After the mid 1980s, they no longer understood ‘programme’ as the central hypothesis of the *C. elegans* project, but as a piece of software which could be designed and run in a minicomputer. The broader versatility of this

apparatus and its more direct relationship with the user led Sulston and other LMB researchers to write their own programs, specifically tailored to the data they needed, and to order to the Laboratory workshop their own computing devices. These devices and programs, in contrast with Brenner's *C. elegans* manifesto, were directed to the 'software' rather than the 'hardware' of the worm*:* they sought the biochemical detail of the nematode's DNA composition rather than the general relationship between its genes and behaviour.

The new role of computers changed the nature and utility of the descriptive data gathered in the project. The mapping data was presented as a means to the genetics of *C. elegans* rather than an end in itself. This made the circulation of data as important as its collection and computer processing. The exchange of mapping information and software within the growing community of worm researchers transformed the way in which the goals of the *C. elegans* project were achieved. Nevertheless, the genetics of development and behaviour were still widely addressed by the worm researchers.

-4.Data circulation and the postponement of goals

The beginning of the mapping initiative is interpreted by Ankeny as a new step in the consolidation of *C. elegans* as a "descriptive model". The mapping and then the sequencing effort multiplied both the amount and range of available data about the worm. This data did not only refer to "genetic sequences", but also to "other biological processes" that could be "correlated with sequences and protein products". The accumulated data is currently

integrated in computer banks and has transformed *C. elegans* into a valuable biomedical "resource". The current importance of the worm database, according to Ankeny, points towards the "need to articulate extensive descriptions of the material to be used before the development of particular hypotheses or theories" (Ankeny, 2001, p. 478; see also Leonelli and Ankeny, this volume).

The overall descriptive orientation of the project is reflected in the connection between the different worm maps. De Chadarevian has shown how in the current computerised WormBase the user can shift, through successive clicks, from Brenner's linkage map to the physical map and the DNA sequence. The different maps, thus, appear "as an integrated series of pictures at increasing resolution". However, the linkage map, the physical map and the sequence relied historically "on different technologies, representational devices, work organisations, institutional set ups and patronage". The linkage map was started "by a single researcher", whereas the physical map represented "the concerted effort of a small group of researchers" and the *C. elegans* sequence "was assembled on industrial lines, in a process involving several hundred people" (de Chadarevian, 2004, p. 108).

These transformations in scale and organisation were, as I will argue, shaped by a simultaneous shift in the nature and use of the descriptive data about *C. elegans.* In Brenner's manifesto papers, data about the worm's mutations, development and behaviour were meant to represent the postulated genetic programme. They were collected through observations and published in voluminous papers which acted as catalogues (e.g. Sulston and Horvitz, 1977; Sulston et al, 1983; White, Brenner et al, 1986). The final aim of the project was to link data about behavioural and developmental patterns with that about their

normal or mutated genetic specification. However, the increasing criticisms to the hypothesis of a genetic programme and the uncertainty of the first results prevented Brenner, White and Sulston from completing the remaining planned descriptions of abnormal cell divisions and neuron connectivity.

The worm's mapping data, by contrast, was generated by minicomputers and stored in tapes or discs. Rather than representing the genetic programme, the data was produced by a computer program which acted as a tool instead of a model of *C. elegans*. The data was, thus, conceived as a means rather than an end: it necessitated the minicomputer user, who reviewed it and decided its future use. This new data regime led Sulston and Coulson to emphasise the "control" they exerted over the mapping software and to circulate, during the construction of the physical map, unpublished data among the international community of *C. elegans* researchers (Sulston et al, 1988, p. 132; Coulson et al, 1986, p. 7825).

This community had considerably grown by the mid 1980s. From 1975 onwards, researchers mainly from the United States became postdoctoral fellows of Brenner, partly attracted by the reputation of the LMB and partly by the worm project. On their return, they established their own *C. elegans* groups and continued working independently on the nematode (Brenner, 2001, chs.7-9; Brown, 2004). Sulston and Coulson's mapping project led the LMB to remain a central node in the international *C. elegans* community (de Chadarevian, 2004, pp. 98 and ff.). This prestige and influence, however, was no longer based on charismatic researchers or attractive hypotheses – such as Brenner's genetic programme – but on the computer programs and data that the new generation of worm workers was both designing and distributing.

An analysis of Coulson's correspondence shows how communication between the LMB and the other worm laboratories was established. From 1985 onwards, he and Sulston received an increasing number of requests to position the genes on which the other groups worked in their ongoing map. The queries were attached to cultures of DNA fragments which either integrated those genes or lay near them. Sulston and Coulson applied to the fragments the fingerprinting technique and compared the results with their mapping database. If one or various overlaps were detected and considered reliable, they forwarded to the laboratories the proposed contig – i.e. a computer printout which represented the delivered DNA fragment and the others with which it overlapped within the worm's genome.[9]

The messages were initially exchanged by post, fax or phone, but in the late 1980s researchers began using ARPANET and BITNET, communication networks based on interconnected computers. This led the number of senders to grow exponentially and the correspondence to become increasingly schematic and impersonalised. The international worm community was, thus, gradually modelling its organisation on expanding computer networks devoted to the delivery and exchange of data. These computer and data components reconfigured the *C. elegans* community, especially with regard to the distribution of work between the growing – and growingly interconnected – worm groups.

In 1988 Robert Waterston, one of Brenner's first postdocs, joined the mapping effort (Sulston and Ferry, 2002, pp. 53-54). This triggered a division of labour in which Waterston led the remaining mapping of *C. elegans*, and Sulston and Coulson focussed on

updating the map and circulating the data. Through internal communication networks, Waterston transmitted the mapping results to the LMB, and Sulston and Coulson incorporated them into the database. The LMB researchers also circulated among the community periodical updates of the whole map.[10]

INSERT FIGURE 3

Both the LMB and the other laboratories benefited from the exchanges. The latter could situate their cultures within the worm's genome and know other adjacent DNA fragments, sometimes candidates to integrate the genes they were chasing. Their next task was normally to sequence the fragments in order to find a nucleotide arrangement characteristic of the genes and their mutations. The mapping scientists refined their contigs by placing the received fragments and, especially, gathered knowledge about the location of genes. This was also achieved by aligning their map with Brenner's previous linkage map, which showed the estimative distances and locations of genes within the worm's chromosomes (see Figure 1). Through this alignment, the loci of Brenner's map could be matched with a series of overlapping DNA fragments which may constitute the genes.

Two main laboratories with which the mapping scientists corresponded during the early stages of the initiative were one led by Samuel Ward at Harvard Medical School and another headed by Robert Horvitz at the Massachusetts Institute of Technology. Both researchers had been postdoctoral fellows of Brenner and, in the late 1970s, Horvitz cooperated with Sulston in the description of the worm's post-embryonic development (Sulston and Horvitz, 1977). An analysis of the mapping requests they submitted shows

that the genes they were chasing were involved in developmental and behavioural mutations of the worm.[11] This suggests a continuation of Brenner's original endeavour by his former postdocs.

The continuation of Brenner's pursuit shows that the mapping initiative, rather than transforming the goals of the *C. elegans* project, redefined the way of achieving them. In other words, the genetic programme, rather than being definitely abandoned was reformulated through new computing technologies, practices and networks. Whereas Brenner directly addressed the problems of development and behaviour and used *C. elegans* genetics to do so, Sulston, Waterston and Coulson allowed other laboratories to address such problems by producing and transmitting computer-based mapping data of the worm's genome. They engaged in a way of conducting research characterised by the postponement of goals and the belief in the power of genetic information to achieve such goals. This implied a reconfiguration of the *C. elegans* community in which Sulston, Coulson and Waterston specialised in the production of data and the other laboratories focussed on its use. This way of proceeding and underlying belief in the potentiality of the gathered data were characteristic of genomics, a field which emerged in the second half of the 1980s and decisively shaped the fate of the worm research.

-5.Genomics and the shift to sequencing

The connection between the *C. elegans* project and other genome-based initiatives started in 1985, with the attendance by worm researchers to the first meetings in which the

feasibility of mapping and sequencing the human genome was debated. De Chadarevian has shown how Sulston, Waterston and Coulson saw in the human genome initiative an opportunity to guarantee the continuity of the worm research. In 1989, James Watson, Director of the newly founded Human Genome Research Institute of the US National Institutes of Health, proposed to Sulston, Coulson and Waterston an agreement by which this institution would partially fund the sequencing of the first three million of the 100 million nucleotide pairs of the worm's genome. The project would be framed in a series of initiatives to assess the technological feasibility of sequencing different organisms before tackling the human genome (de Chadarevian, 2004, pp. 103 and ff.).

This patronage led the *C. elegans* researchers to subsequently frame their work in the proposed sequencing of the human genome. Sulston, Waterston and Coulson presented the worm sequencing project as "a technological pilot for the human genome (which is some 40 times larger)" in both scientific publications and applications for further support. The nematode's sequence was defined as an "essential" contribution to the "understanding" of the "biology of man" (Sulston et al, 1992, p. 37).[12] In 1993, Sulston became Director of the Sanger Centre, a new institute in Cambridgeshire supported by the MRC and the emerging biomedical charity Wellcome Trust. It was conceived to host the British participation in the Human Genome Project (HGP), which had been officially launched three years before as an international effort aimed to coordinate previous local initiatives.

The Sanger Centre hosted a sequencing initiative which was performed "on an industrial scale" and directed to both worm and human genome. The physical map of *C. elegans* was also finished in its laboratories, despite this project receiving less publicity than the

sequencing initiatives. The sequence of *C. elegans* was completed in 1998 by an "international consortium" which included the Sanger Centre, Waterston's laboratory in Washington University-St. Louis and other members of the international worm community. Two years later, the first draft of the human genome – in which both the Sanger Centre and Waterston's laboratory were heavily involved – was announced (de Chadarevian, 2004, pp. 105 and 107).

The connections between *C. elegans* and human genome sequencing have led the worm project to be seen as a *precursor* of the HGP. This association is especially strong in the narratives on the "origins" of the HGP, which proliferated after the announcement of the Project's first draft (2000). Historian and sociologist Michael Fortun has shown the difficulties of associating the HGP with a single entity. He rather refers to the "genomics project" and identifies a series of local mapping and sequencing initiatives, which were driven by diverse motivations and gradually coalesced in the new field of genomics during the late 1980s (Fortun, 1993, 1999).[13]

In the case of *C. elegans*, the association with the HGP secured the continuity of the project, but also shaped inevitably its trajectory. Sulston, Waterston and Coulson needed to embark on sequencing the worm and scale-up to humans, despite these objectives not being originally contemplated, even after the launch of the physical map of *C. elegans*. Other long-term mapping initiatives of model organisms on yeast and the bacterium *E. coli* (García-Sancho, 2008, pp. 117 and ff.) were equally reconstructed as *pilots* for human sequencing and incorporated to the logic of progress which defined the narratives on the origins of the HGP.

This generalised shift to large-scale sequencing triggered disciplinary reconfigurations. In 1987, the journal *Genomics* was launched as the publication medium of a "new discipline" engaged with mapping and sequencing whole genomes of different organisms, including humans. The journal's first editorial linked the resulting DNA maps and sequences to promises and expectations: the data would be the "rosetta stone" from which "the complexities of gene expression in development" could be "translated and the genetic mechanisms of disease interpreted" (McKusick and Ruddle, 1987, p. 1; Powell et al, 2007). These promises were largely believed by science policy agencies, entrepreneurs, biomedical funders and the general public at a time in which the perception of them living in an "information society" was spreading. In this information society, computer-assisted control and access to data – economic, biotechnological or of any other kind – were considered to be the main means of increasing productivity, knowledge and welfare (Kline, 2006; García-Sancho, 2009).

The scientific and socio-political concern with information led the *C. elegans* community to strengthen the data-intensive dimension of its research. From the late 1980s onwards, the worm project was increasingly presented as a data-driven endeavour, led by the collection and analysis of information rather than by the traditional formulation of hypotheses (e.g. Sulston and Ferry, 2002, pp. 58-59). Computing technologies such as the nematode database (Leonelli and Ankeny, this volume) were emphasised within a general view of biomedicine as being transformed into an information science (Lenoir, 1999; Gilbert, 1992).

This emphasis resulted in an increasing abstraction of the original DNA molecule. The DNA samples that researchers attached to their *C. elegans* mapping requests were gradually substituted by the exclusive circulation of mapping and sequencing data. As a consequence of this, the use of the data – and not only the goals of the worm project – was increasingly postponed. The worm researchers focussed on fulfilling the mapping and sequencing goals of their new funders, without directly addressing what they were going to do with the resulting information. The main claim about the applicability of this information was that it would contribute to the interpretation of the "human sequence" (e.g. Sulston et al, 1992, p. 37).

By the closure of the 20th century, the *C. elegans* project had, thus, been reformulated according to the *data-intensive* and *dry biology* that the spokespersons of genomics were proposing as new and promising ways of conducting science. However, the history of the investigations on this tiny worm shows that the emphasis on data and computation were present since the very beginning of the project. What had been rather transformed 30 years after Brenner's original proposal were the meanings and practices that the *C. elegans* researchers attached to the notions of 'data' and 'computation'. Understanding these historical transformations help to overcome propagandistic discourses and grasp what is behind the proposal and promises of a *data-driven* science.

-6.Conclusion

This paper has shown how the history of the computing technologies and practices adopted by the researchers working on the worm *C. elegans* help to better characterise the process by which this model organism was reconfigured as a privileged object of the newly proposed data and computer-intensive biology. From the mid 1980s onwards, the worm shifted from being defined as a model of the genetic basis of development and behaviour to become a platform to assess the feasibility of sequencing of the human genome. This transition had a transforming effect on the identity of the investigations on *C. elegans*, which after the determination of the human genome sequence (2000) were presented as an *antecedent* of this large-scale data-gathering initiative. The sequencing of both human and worm genomes has been subsequently considered an exemplar of the new data-driven research, founded on big science efforts to collect, store and deliver information which is believed to provide fundamental future scientific insights.

The evolving place of the computer and of the practice of computation in *C. elegans* research demonstrates that even in contemporary and cutting-edge scientific fields, a historical perspective is necessary to adequately understand and assess the transformations that the new field has fostered. My account complements previous philosophical and historical investigations on *C. elegans* and specifically links them to the claims of novelty attached to the proposed data-driven biology. Its main argument is that only by comparing the computers that were available before and after the worm mapping and sequencing initiative, and the way researchers used them, it is possible to fully grasp the meanings and expectations currently attributed to the computation of *C. elegans* and other biomedical data.

With the advent of the worm mapping initiative, the researchers involved shifted from a conceptualisation of the genetics of *C. elegans* in a mainframe apparatus to writing minicomputer programs which automatically provided map and then sequence data of the nematode's genes. This shift redefined the long-standing aim of describing the worm and the role data played in such a description. Data was no longer evidence to be used in order to test a pre-existing hypothesis on the genetic programme of *C. elegans*. Rather, data was regarded as a computer output which was distributed among the laboratories working on the genetics of the worm. This led the hypothesis of the genetic programme to subsequently evolve hand-to-hand with the production of mapping and sequencing software, and with the biological significance that the data receivers attached to such information. Thus, data became key to shaping and guiding the research hypothesis behind the *C. elegans* project and the organisation of its growing international community.

The historicity of 'computer', 'computing', 'data' and 'programme' in the *C. elegans* project also reflects the contributions that the history of computing can offer towards the writing of a renewed 20th century history of biology. The integration of both historical traditions may lead to accounts which will not be centred only on biological concepts or hypotheses – such as the 'genetic programme' of *C. elegans* – but also, and increasingly, on computational practices and instruments. Computing instruments are not sufficient to change the course of a scientific project, as shown by Sydney Brenner's pursuit of the worm's genetic programme, which resisted the introduction of minicomputers into his group and their use by John White. However, when these instruments are associated to a practice – such as map and sequence assemblage – and this practice becomes the main goal

of the investigations – i.e. the construction of a physical map of *C. elegans* – the aims and hypothesis behind the project are significantly reformulated.

The accounts of the computerisation of biology need, thus, to distinguish between computers as conceptual models – as in Brenner's genetic programme – as tools at the service of pre-existing hypotheses – as in White's neuron connectivity programs – or as embodiments of such hypotheses – as in the mapping and sequencing software. When the genetic programme of *C. elegans* became embodied in a computer program, the histories of both entities coalesced and evolved together, forming a privileged field for the unification of the historiographies of biology and computing.[14] This coalescence was fostered by a historically specific computational technology: the minicomputer, in which researchers had access not only to the results, but to their production by the operation of internal software. In this regard, the novelty of *C. elegans* research as a data-driven project was not so much the application of computers, but the end of the separation between the computer and the user biologist. This confluence, paradoxically, led to an increasing gulf between the sample DNA fragments, the map and sequence data derived from them, and the future use of such data.

The convergence between computing instruments and biomedical practice does not imply that bioinformatics software and genomic technologies are hypothesis-free. The mapping and sequencing of *C. elegans* was guided by the belief in the potentiality of DNA data to deduce how genes worked. This belief in the power of information was the centripetal force behind the convergence and has become the central – and often overlooked – hypothesis of data-driven science. *C. elegans* mapping and sequencing data was not initially gathered for

its own sake, but to be distributed among laboratories working on the long-lasting objectives of the worm project. The data, thus, represented an alternative strategy for determining the genetic programme, i.e. the relationship between the nematode's genes and its development and behaviour.

By the time the *C. elegans* project became a data-intensive endeavour, it was difficult to escape from the belief in the power of information. The rise and increasing favour of genomics led the producers of the worm map to engage in a large-scale sequencing initiative of both nematode and human in order to guarantee the survival of the project. This provided the accumulating data with an own life, in a journey which was as inevitable as unpredictable. In his memoirs, John Sulston recalls having heard "the prison door shut" behind them when he and his collaborators accepted the worm sequencing project (Sulston and Ferry, 2002, p. 13). After its conclusion, the next uncertainty in this long one-way road is what to do with the gathered data.

ACKNOWLEDGEMENTS

Special thanks are given to the Editor of this special issue, Sabina Leonelli, for her patience, understanding and support. My PhD Supervisor, Andrew Mendelsohn, provided insightful comments on an early draft of this manuscript, as well as John Sulston, Alan Coulson and two anonymous referees. My current Line Manager, María Jesús Santesmases, has also provided invaluable feedback and granted me with time to write and revise this paper.

The Genetics Society of America, Oxford University Press, Elsevier, the Science Museum in London, the Computer History Museum, the National Archives of the UK, and the Archives of the Medical Research Council, Wellcome Trust and MRC Laboratory of Molecular Biology of Cambridge (LMB) allowed me to reproduce copyrighted and unpublished material. The archivist of the LMB, Annette Faux, provided invaluable help in searching particularly difficult records.

The research on which this paper is based was conducted while holding doctoral fellowships awarded by Caja Madrid Foundation, Madrid City Hall and Residencia de Estudiantes (Spain), together with a Hans Rausing Fellowship awarded by the Centre for the History of Science, Imperial College, London. Fieldwork and other expenses were covered by small grants awarded by the Royal Historical Society. Manuscript preparation was made possible by a postdoctoral contract awarded by the Spanish National Research Council (CSIC) and the projects FFI2009-07522 and HUM2006-04939/FISO, funded by the Spanish Ministry of Science and Innovation.

NOTES

[1] On the impact of cybernetics on genetics see Kay, 2000; Sarkar, 1996; Fox-Keller, 1995; Brandt, 2005; García-Sancho, 2007ab. Historians are beginning to document the instruments, practices and algorithms in which this information discourse materialised in genetics and other fields of biology during the second half of the 20th century (November, 2006; Suárez-Díaz and Anaya-Muñoz, 2008; Suárez-Díaz, 2010; Strasser, 2010; García-Sancho, 2011; Chow-White and García-Sancho, in press).

[2] The close connection between nematode worms and Brenner's approach suggests that, as Greg Mitman and Anne Fausto-Sterling have argued, model organisms are used to defend a particular perspective on biological research. Their implementation, therefore, largely depends on the acceptance of such perspective (Mitman and Fausto-Sterling, 1992; see also Geison and Laubichler, 2001, pp. 12 and ff.; Burian, 1993).

[3] This descriptive orientation in *C. elegans* research has been seen by de Chadarevian as evidence that "much of the work of molecular biologists in the study of development – and in 'classical' molecular biology in general – was of the analytical / comparative rather than of the experimental kind most commonly associated with the molecular paradigm" (de Chadarevian, 2000, p. 381). The transparent body of *C. elegans* crucially eased the observation and description of its cells during development. Richard Doyle has identified this transparency with "the postvital organism", where "interior and exterior, genetics and anatomy, implode under the gaze and touch of research" (Doyle, 1997, p. 14).

[4] The MRC does not seem to have had a specific programme to foster the introduction of computers into biomedical laboratories, as was the case of its US counterpart, the National Institutes of Health (November, 2006, ch.3). Applications for the purchase of minicomputers were similar in their format to requests for research funds and always signed by high-profile LMB faculty members (S. Brenner, 1985: "Future in-house computing at LMB". In Archives of the MRC Laboratory of Molecular Biology, Cambridge, UK. Uncatalogued file on Computers).

[5] String manipulation algorithms had been used since the 1950s by computer scientists to edit and find patterns in a text with the instructions of a particular mainframe or minicomputer program (Haigh, 2006, pp. 6-13). In the subsequent decade, those same algorithms were widely applied to database searches – including databases with scientific literature – and to the analysis of protein sequences (Strasser, 2010; Suárez-Díaz and Anaya-Muñoz, 2008; Suárez-Díaz, 2010).

[6] S. Brenner, 1985: Letter to the MRC Head Office (Archives of the MRC Laboratory of Molecular Biology, Cambridge, UK. Uncatalogued file on Computers). The minicomputer was also requested to process microscopic images of cells produced at the LMB and to build a database with the accumulated neuron connectivity data on *C. elegans*.

[7] This comparative approach to determine the position of the bands was inspired by the Viterbi algorithm, a variation of which was incorporated to the mapping programs. The Viterbi algorithm was used in the then emerging speech recognition software, which transformed spoken language into written text by reference to a pre-recorded set of sounds (Sulston et al, 1988, pp. 126-27). The scanning densitometer was custom-built despite a commercial version being available via the company Amersham International (Gee, 2007, ch.4). In order to plan its manufacturing, the *C. elegans* team visited the Clinical and Population Cytogenetics Unit in Edinburgh, which was using a similar scanning device to process chromosomal images (Sulston et al, 1988, pp. 125 and 132; id., 1989, p. 106; see also de Chadarevian, 2009).

[8] This defence of human judgement contrasted with the fully automated approaches which were then spreading in DNA sequencing (García-Sancho, forthcoming, chs.5-6). The *C. elegans* team stated among the advantages of human intervention its "low capital cost" and the "provision of a high resolution image for direct comparison" with the autoradiographs. These possibilities needed to be "weighted against the advantages of more fully automatic methods" (Sulston et al, 1989, p. 106).

[9] Papers and Correspondence of A. Coulson, donated by the Archives of the MRC Laboratory of Molecular Biology of Cambridge and currently being catalogued at the Archives of the Wellcome Library, London, UK. Collection reference PP/COU.

[10] Sulston, Coulson and Waterston were committed to the free circulation of mapping information and software, and made of it a sign of identity within the *C. elegans* community. In this regard, they shared the

rhetoric of cooperation and the "moral economy" of free exchange that Robert Kohler has identified in the *Drosophila* community during the early and mid 20th century (Kohler, 1994, pp. 133-135). However, whereas the *drosophilists* exchanged mutants and know-how, genetic information became the main outcome to be circulated during the worm's mapping project. This suggests that the traditional exchange of specimens in the life sciences was being gradually substituted by the exchange of data (García-Sancho, 2008, pp. 140-46 and 239-44; id., 2009). On the importance of moral economies of exchange and economies of bio-data see Strasser, 2010; Parry, 2004; Harvey and McMeekin, 2007; Lewis, 2010.

[11] Letter of Junying Yuan – Horvitz's fellow – to Sulston (1985) and letter of Ward to Coulson (1987). Papers and Correspondence of A. Coulson, donated by the Archives of the MRC Laboratory of Molecular Biology of Cambridge and currently being catalogued at the Archives of the Wellcome Library, London, UK. Collection number PP/COU.

[12] J. Sulston and A. Coulson (1989) "Mapping and sequencing the genome of *Caenorhabditis elegans.* Application for a project grant". In Archives of the Medical Research Council, London (UK), file on Human Genome Mapping Project (reproduced with the permission of the MRC following a Freedom of Information Request submitted in 2006). In their application, Sulston and Coulson requested to acquire two computers, one of them a Microvax II, also manufactured by DEC and chosen for software compatibility and networking with the existing LMB computer.

[13] The British local initiative was the Human Genome Mapping Project (HGMP), which was established by the MRC in the late 1980s following recommendation of Brenner, then Director of a laboratory outside the LMB. It was intended as a concerted UK effort in which the experience of Brenner's former LMB colleagues "with the nematode genome" would provide "a useful benchmark" for assessing "the scale" of mapping human DNA (S. Brenner, 1986: "A physical map of the human genome", p. 2, and id., undated: "Map of man". In The National Archives of the UK, Kew, document reference FD 23/3441). In 1990, the HGMP was incorporated to the Human Genome Project.

[14] Biomedical databases, sequence-analysis algorithms and computer-operated automatic sequencers are other hybrid instruments which allow a combination of the history of biology and computing. For historical research on these technologies see Suárez-Díaz and Anaya Muñoz, 2008; Strasser, 2010; November, 2006; Chow-White and García-Sancho, in press; García-Sancho, forthcoming, Parts II and III.

REFERENCES

Ankeny R. (1997). The Conqueror Worm: A Historical and Philosophical Examination of the Use of the Nematode *C. elegans* as a Model Organism. PhD Dissertation: University of Pittsburgh.

Ankeny R. (2000). Fashioning descriptive models in biology: of worms and wiring diagrams. Philosophy of Science, 67 Supplement, S260-S272.

Ankeny R. (2001). The natural history of *C. elegans* research. Nature Review Genetics, 2, 474-479

Ankeny R. (2006). Wormy logic: model organisms as case-based reasoning. London School of Economics: Working Papers on the Project 'The Nature of Evidence: How Well Do Facts Travel?' (preprint number 07/06).

Ankeny (2010). Historiographical reflections on model organisms: or, how the mureaucracy may be limiting our understanding of contemporary genetics and genomics. History and Philosophy of the Life Sciences, 32(1), 91-104.

Bennet, J.M. and Kendrew, J. (1952). The computation of Fourier syntheses with a digital electronic calculating machine. Acta Crystallographica, 5, 109-116.

Brandt C. (2005). Genetic code, text, and scripture: Metaphors and narration in German molecular biology. Science in Context, 18(4), 629-648.

Brenner S. (1963a). Letter to Max Perutz. Reprinted in W. Wood (ed., 1988) The nematode *Caenorhabditis elegans* (pp. X-XI). New York: Cold Spring Harbor.

Brenner, S. (1963b). Excerpts from proposal to the Medical Research Council. Reprinted in W. Wood (ed., 1988) The nematode *Caenorhabditis elegans* (pp. XI-XIII). New York: Cold Spring Harbor.

Brenner S. (1973). The genetics of behaviour. British Medical Bulletin, 29(3), 269-271.

Brenner, S. (1974a). New directions in molecular biology. Nature, 248, 785-787.

Brenner S. (1974b). The genetics of *Caenorhabditis elegans*. Genetics, 77, 71-94.

Brenner S. (2001). My Life in Science. London: BioMed.

Brown A. (2004). In the Beginning was the Worm: Finding the Secrets of Life in a Tiny Hermaphrodite. New York: Columbia University Press.

C. elegans Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans:* a platform for investigating biology. Science, 282, 2012-2018.

Chow-White P. and García-Sancho M. (in press). Bi-directional shaping and spaces of convergence: interactions of biology and computing from the first DNA sequencers to global genome databases. Science, Technology & Human Values. http://dx.doi.org/10.1177/0162243910397969

Coulson A., Sulston J., Brenner S. and Karn J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proceedings of the National Academy of Sciences of the US, 83, 7821-7825.

De Chadarevian S. (1998). Of worms and programmes: *Caenorhabditis elegans* and the study of development. Studies in History and Philosophy of Biological and Biomedical Sciences, 29, 81-105.

De Chadarevian S. (2000). Mapping development or how molecular is molecular biology? History and Philosophy of the Life Sciences, 22(3), 381-396.

De Chadarevian S. (2002). Designs for Life: Molecular Biology after World War II. Cambridge: Cambridge University Press.

De Chadarevian S. (2004). Mapping the worm's genome: tools, networks, patronage. In H.J. Rheinberger and J.P. Gaudillière (eds.) From Molecular Genetics to Genomics: The Mapping Cultures of Twentieth Century Biology (pp. 95-110). London and New

York: Routledge.

De Chadarevian S. (2009). Viewing chromosomes. In S. Brauckmann, C. Brandt, D. Thieffry and G.B. Müller (eds.) Graphing Genes, Cells and Embryos (pp. 57-62). Berlin: Max Planck Institute for the History of Science, preprint number 380.

Doyle R. (1997). On Beyond Living: Rhetorical Transformations of the Life Sciences. Stanford: Stanford University Press.

Fortun M. (1993). Mapping and Making Genes and Histories: The Genomics Project in the United States, 1980-1990. PhD Dissertation: Harvard University.

Fortun M. (1999). Projecting speed genomics. In id. and E. Mendelsohn (eds.) The Practices of Human Genetics (pp. 25-48). Dordrecht: Kluwer.

Fox Keller E. (1995). The body of a new machine. In id., Refiguring Life: Metaphors of Twentieth Century Biology (pp. 79-118). New York: Columbia University Press.

García-Sancho M. (2007a). The rise and fall of the idea of genetic information (1948-2006). Genomics, Society and Policy, 2(3), 16-36.

García-Sancho M. (2007b). Mapping and sequencing information: the social context for the genomics revolution. Endeavour, 31(1), 18-23.

http://dx.doi.org/10.1016/j.endeavour.2007.01.006

García-Sancho M. (2008). Sequencing as a Way of Work: A History of its Emergence and Mechanisation – From Proteins to DNA, 1945-2000. PhD Dissertation: Imperial College, London.

García-Sancho M. (2009). The perception of an information society and the emergence of the first computerised biological databases. In A. Matsumoto and N. Nakano (eds.) Human Genome. Features, Variations and Genetic Disorders (pp. 257-276). New York: Nova Science Publishers.

García-Sancho M. (2010). A new insight into Sanger's development of sequencing: from proteins to DNA, 1943-1977. Journal of the History of Biology, 43(2), 265-323. http://dx.doi.org/10.1007/s10739-009-9184-1

García-Sancho M. (2011). From metaphor to practices: the introduction of information engineers into the first DNA sequence database. History and Philosophy of the Life Sciences, 33(1): 71-104.

García-Sancho (forthcoming). Biology, Computing and the History of Molecular Sequencing: From Proteins to DNA, 1945-2000. Basingstoke: Palgrave Macmillan.

Gee S. (2007). Case study of Amersham and automated DNA sequencing instruments. In id., The Case of Research Tools for Drug Discovery: Evolving User-Producer Inter-Dependencies and the Exchange of Knowledge. PhD Dissertation: University of Manchester.

Geison G.L. and Laubichler M.D. (2001). The varied lives of organisms: variation in the historiography of the biological sciences. Studies in History and Philosophy of Biological and Biomedical Sciences, 32, 1-29.

Gilbert W. (1992). A vision of the grail. In D.J. Kelves and L. Hood (eds.) The Code of Codes: Scientific and Social Issues in the Human Genome Project (pp. 83-97). Cambridge: Harvard University Press.

Haigh T. (2006). Remembering the office of the future: the origins of word processing and office automation. Annals of the History of Computing, 28(4), 6-31.

Kay L. (2000). Who Wrote the Book of Life: A History of the Genetic Code. Stanford: Stanford University Press.

Kline R. (2006). Cybernetics, management science and technology policy: the emergence of 'information technology' as a keyword, 1948-1985. Technology and Culture, 47, 513-535.

Kohler R. (1994) *Lords of the Fly: Drosophila Genetics and the Experimental Life*. Chicago: University of Chicago Press.

Burian R. (1993). How of the choice of experimental organism matters: epistemological reflections on an aspect of biological practice. Journal of the History of Biology 26(2), 351-367 (special section on The Right Organism for the Job).

Lenoir T. (1999). Shaping biomedicine as an information science. In M.E. Bowden, T.B. Hahn and R.V. Williams (eds.) Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems (pp. 27-45). Medford: ASIS.

Leonelli S. (2007a). Weed for Thought: Using Arabidopsis Thaliana to Understand Plant Biology. PhD Dissertation: Vrije Universiteit, Amsterdam.

Leonelli S. (2007b). Growing weed, producing knowledge: an epistemic history of *Arabidopsis thaliana*. History and Philosophy of the Life Sciences, 29, 193-224.

Leonelli (2010a). Documenting the emergence of bio-ontologies: or, why researching bioinformatics requires HPSSB. History and Philosophy of the Life Sciences, 32, 105-125.

Leonelli (2010b). Packaging small facts for reuse: databases in model organism biology. In P. Howlett and M. Morgan (eds.) How Well Do Facts Travel? The Dissemination of Reliable Knowledge (pp. 325-348). Cambridge: Cambridge University Press

Lewin, R. (1984). Why is development so illogical? Science, 224, 1327-1329.

Lewis, J. (2010). Matchmaking mechanisms: collaborative arrangements in proteomics and bioinformatics". In J.N. Parker, N. Vermeulen and B. Penders (eds.) Collaboration in the New Life Sciences (pp. 179-199). London: Ashgate.

Löwy I. and Gaulillière J.P. (1998). Disciplining cancer: mice and the practice of genetic purity. In ids. (eds.) The Invisible Industrialist: Manufacturers and the Construction of Scientific Knowledge (pp. 209-249). Basingstoke: Palgrave Macmillan.

McKusick, V. and Ruddle, F. (1987). Editorial: a new discipline, a new name, a new journal. Genomics, 1, 1-2.

Mitman G. and Fausto-Sterling A. (1992) Whatever happened to *Planaria*? C.M. Child and the physiology of inheritance. In A.E. Clarke and J.H. Fujimura (eds.) The Right Tools for the Job: At Work in Twentieth Century Life Sciences (pp. 172-196). Princeton: Princeton University Press.

Moody, G. (2004). Digital Code of Life: How Bioinformatics is Revolutionising Science, Medicine and Business. London: Wiley.

Morange M. (1997). The transformation of molecular biology on contact with higher organisms, 1960-1980: from a molecular description to a molecular explanation. History and Philosophy of the Life Sciences, 19, 369-393.

Powell A., O'Malley M., Müller-Wille S., Calvert J., Dupré J. (2007). Disciplinary baptisms: a comparison of the naming stories of genetics, molecular biology, genomics and systems biology. History and Philosophy of the Life Sciences, 29, 5-32.

Rader K. (2004). Making Mice: Standardising Animals for American Biomedical Research, 1900-1955. Princeton: Princeton University Press.

Sanger F. and Coulson A. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology, 94, 441-448.

Sanger F., Nicklen S. and Coulson A. (1977). DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Sciences of the US, 74, 5463-5467.

Sarkar S. (1996). Biological information: a sceptical look at some central dogmas in molecular biology. In id. (ed.) The Philosophy and History of Molecular Biology: New Perspectives (pp. 187-233). Dordrecht: Kluwer. Reprinted in id. (2005). Molecular

Models of Life: Philosophical Papers on Molecular Biology (pp. 205-260). Cambridge: MIT.

Staden, R. (1984a) Computer Methods to Aid in the Determination and Analysis of Nucleic Acid Sequences. PhD Dissertation: University of Cambridge.

Staden R. (1984b). A computer program to enter DNA gel reading data into a computer. Nucleic Acids Research, 12, 499-503.

Strasser B. (2010). Collecting, comparing, and computing sequences: the making of Margaret O. Dayhoff's 'Atlas of Protein Sequence and Structure', 1954–1965. Journal of the History of Biology, 43, 623-660.

Suárez-Díaz E. (2010). Making room for new faces: evolution, genomics and the growth of bioinformatics. History and Philosophy of the Life Sciences, 32, 65-90.

Suárez-Díaz E. and Anaya-Muñoz V. (2008). History, objectivity and the construction of molecular phylogenies. Studies in History and Philosophy of Biological and Biomedical Sciences, 39 , 451-468.

Sulston J. and Horvitz, R. (1977). Post-embryonic cell lineages of the nematode *Caenorhabditis elegans*. Developmental Biology, 56, 110-156.

Sulston J., Schierenberg E., White J. and Thompson J. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Developmental Biology, 100, 64-119.

Sulston J., Mallett F., Staden R., Durbin R., Horsnell T. and Coulson A. (1988). Software for genome mapping by fingerprinting techniques. Computer Applications in Biosciences, 4 (1), 125-132.

Sulston J., Mallett F., Durbin R. and Horsnell T. (1989). Image analysis of restriction enzyme fingerprint autoradiograms. Computer Applications in Biosciences, 5 (2), 101-106.

Sulston J. Du Z., Thomas K., Wilson R., Hillier L., Staden R., Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A., Craxton M., Durbin R., Berks M., Metzstein M., Hawkins T., Ainscough, R., Waterston R. (1992). The *C. elegans* genome sequencing project: a beginning. Nature, 356, 37-41.

Sulston J. and Ferry G. (2002). The Common Thread: A Story of Science, Politics, Ethics and the Humane Genome. London: Bantam.

Wieber F. (2006). Interplay between molecular biology and computational chemistry: models and simulations in the construction of a computational protein chemistry (1960-1980). In S. de Chadarevian and H.J. Rheinberger (eds.) History and Epistemology of

Molecular Biology and Beyond: Problems and Perspectives (pp. 95-103). Berlin: Max Planck Institute for the History of Science, preprint number 310.

White, J. (1974) Computer Aided Reconstruction of the Nervous System of Caenorhabditis elegans. PhD Dissertation: University of Cambridge.

White J., Southgate E., Thompson J. and Brenner S. (1986) The structure of the nervous system of the nematode *Caenorhabditis elegans* (AKA The mind of a worm). Philosophical Transactions of the Royal Society of London, B series, 314, 1-340.

Yi D. (2008) Cancer, viruses and mass migration: Paul Berg's venture into eukaryotic biology and the advent of recombinant DNA research and technology, 1967-1980. Journal of the History of Biology, 41, 589-636.