

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

DNA as information: At the crossroads between biology, mathematics, physics and chemistry

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

DNA as information: At the crossroads between biology, mathematics, physics and chemistry / Cartwright, Julyan H. E.; Giannerini, Simone; Gonzalez, DIEGO LUIS. - In: PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES A: MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES. - ISSN 1364-503X. - STAMPA. - 374:2063(2016), pp. 20150071.1-20150071.9. [10.1098/rsta.2015.0071]

Availability:

This version is available at: <https://hdl.handle.net/11585/570929> since: 2016-11-27

Published:

DOI: <http://doi.org/10.1098/rsta.2015.0071>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Cartwright Julyan H. E., Giannerini Simone and González Diego L. 2016. DNA as information: at the crossroads between biology, mathematics, physics and chemistry. *Phil. Trans. R. Soc. A.* 374: 20150071.

<http://doi.org/10.1098/rsta.2015.0071>

The final published version is available online at:

<http://doi.org/10.1098/rsta.2015.0071>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.



DNA as Information: at the Crossroads Between Biology, Mathematics, Physics, and Chemistry

Julyan H. E. Cartwright^{1,2}, Simone
Giannerini³ and Diego L. González^{3,4}

¹Instituto Andaluz de Ciencias de la Tierra,
CSIC–Universidad de Granada, E-18100 Armilla,
Granada, Spain

²Instituto Carlos I de Física Teórica y Computacional,
Universidad de Granada, E-18071 Granada, Spain

³Dipartimento di Scienze Statistiche “Paolo Fortunati”,
Università di Bologna, I-40126 Bologna, Italy

⁴Istituto IMM-CNR, Area della Ricerca CNR di
Bologna, I-40129, Bologna, Italy

On the one hand, biology, chemistry and also physics tell us how the process of translating the genetic information into life could possibly work, but we are still very far from a complete understanding of this process. On the other hand, mathematics and statistics give us methods to describe such natural systems — or parts of them — within a theoretical framework. Also, they provide us with hints and predictions that can be tested at the experimental level. Furthermore, there are peculiar aspects of the management of genetic information that are intimately related to information theory and communication theory. This theme issue is aimed at fostering the discussion on the problem of genetic coding and information through the presentation of different innovative point of views. The aim of the editors is to stimulate discussions and scientific exchange that will lead to new research on why and how life can exist from the point of view of the coding and decoding of genetic information. The present introduction represents the point of view of the Editors on the main aspects that could be subject of the future scientific debate.



Figure 1. The Graces, Asclepius, and Mercury, a classical Roman bas-relief now in the Vatican Museum in Rome, depicted in an engraving by Giacomo Bossi published in 1788 [2]. The bas-relief shows a bearded man with one knee on the ground and his hands raised giving thanks to Asclepius, presumably for an important act of healing; recall that Asclepius symbolizes medicine. However, if we are allowed a free interpretation of the image, the analogy with the synthesis of proteins is impressive². We see Mercury with his helmet and caduceus. The latter is formed by two serpents intertwined on a rod in the fashion of the double helix of DNA. The bearded man may be interpreted as an envoy of Mercury to Asclepius; it seems that Mercury is pushing the man towards Asclepius. Asclepius is symbolized by a serpent coiled on a rod in analogy with the single helix of RNA. These two subjects together represent symbolically the second step in the synthesis of proteins, the formation of messenger RNA using the information that arises from DNA. Finally we have the three Graces. Going back to their Greek origins we find that they induce in Gods and mortals the joy and beauty of nature. Proteins are plural — three Graces and not a single God — and are the fundamental agents of metabolism and, as such, are responsible for all the qualities that we imagine associated with life. With little effort we can consider proteins as the three Graces of life, thus completing in an admirable way the analogy between DNA and information and this beautiful Roman bas-relief that we have also put on the cover of this issue of Philosophical Transactions.

“Natural History may be considered in two very different lights: either, 1st, as a collection of facts and objects presented by nature, from the examination, analysis, and combination of which we acquire whatever knowledge we are capable of attaining both of the order of nature, and of the agents she employs for producing her ends, and from which, therefore, all sciences arise; or, 2dly, as an assemblage of phenomena to be explained; of effects to be deduced from causes; and of materials prepared to our hands, for the application of our principles to useful purposes. Natural history, therefore, considered in the one or the other of these points of view, is either the beginning or the end of physical science.”

John Herschel, *A Preliminary Discourse on the Study of Natural Philosophy*, 1830 [1]

1. Biological complexity

²The so-called central dogma of molecular biology was enunciated by Francis Crick in the 1950s [3] and proposes that genetic information can be transmitted from nucleic acids to proteins and not in the reverse way: “The central dogma of molecular

Biology cannot escape from the laws of physics and of chemistry; yet biology differs from physics and chemistry in that living beings show behaviour which, while compatible with these laws, is not observed in other systems; for this reason we call physical and chemical systems inanimate, as they do not display those qualities inherent in life. A key question then is to explain the particular behaviour of living beings; why life is rather more than the sum of its parts. This is not to subscribe to vitalism, but to note that life has the capacity to operate and adapt in conditions of constant change; in a dynamical environment. This capacity to create order, to decrease entropy — at the cost of greater disorder somewhere else, since biology operates within the laws of thermodynamics — is a fundamental aspect of life. What then is life? Many scientists have provided definitions, all of which have difficulties with some particular instance or other. For our present purposes we argue that an exact definition of life is not so important; the notorious definition of pornography — “I know it when I see it” (United States Supreme Court Justice Potter Stewart’s opinion in the case of *Jacobellis versus Ohio* (1964) [6]) — suffices, because we are not so much interested in attempting a binary classification, into the living and non-living, as in quantifying what it is that makes biological systems special.

The complexity present in a tiny amount of living matter is staggering; a single human cell is a veritable factory containing many specialized elements and possessing about 2 metres of genetic instructions in the form of DNA. This complexity has arisen as living things have passed through the filter of natural selection. If we look back to the beginnings of life, there there must have existed a sort of porous wall in time before which there were only molecules subject to the laws of thermodynamics and mechanics but after which there were complex self-organizing, self-assembling, self-reproducing and self-adapting systems of molecules, now subject to evolution. Whatever may be the point or points in space and time at which this barrier was passed — however and wherever life arose — it must have been passed at some point, and understanding the passage must be vital to arriving at a theoretical biology that may be called a theory of life. As Sidney Brenner has put it [7]:

In one sense, everything in biology has already been ‘published’ in the form of DNA sequences of genomes; but, of course, this is written in a language we do not yet understand. Indeed, I would assert that the prime task of biology is to learn and understand this language so that we could then compute organisms from their DNA sequences. . . . We are at the dawn of proper theoretical biology.

It is our view that it is this emergence of complex interacting elements that are acting ultimately under the laws of physics and chemistry that makes living matter different from the inanimate. However, the complexity attained depends on a system’s history; on its evolution. While the involvement of the past of a system in its present can be found in some physical and chemical systems, in phenomena ranging from hysteresis on short timescales to geological pattern formation at long times, the significant difference in living matter is the accompanying growth of complexity (it is worthy of note that biology shares this trait with technological and economic systems, as well as with culture and the arts).

We now have the human genome and we find *Homo sapiens* possesses some 23 000 genes [8]. A nematode worm, *Caenorhabditis elegans*, on the other hand, has 19 000 [9]. Why does a ‘lowly’ worm need so many genes, while we humans, apparently so much more complex — or, at least, we should like to think so —, have hardly many more? Are we really little more than worms? (It

biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid” [4]. Crick later wrote about his choice of term: “I called this idea the central dogma, for two reasons, I suspect. I had already used the obvious word hypothesis in the sequence hypothesis, and in addition I wanted to suggest that this new assumption was more central and more powerful. . . . As it turned out, the use of the word dogma caused almost more trouble than it was worth. Many years later Jacques Monod pointed out to me that I did not appear to understand the correct use of the word dogma, which is a belief that cannot be doubted. I did apprehend this in a vague sort of way but since I thought that all religious beliefs were without foundation, I used the word the way I myself thought about it, not as most of the world does, and simply applied it to a grand hypothesis that, however plausible, had little direct experimental support.” [5]

gets worse, some plants have enormous genomes: *Paris japonica*'s is about 150 billion base pairs long, some 50 times the human genome [10].) We can see two complementary answers to these riddles: the first is that the genome, the information used to construct a man or a worm, which is written in language we have yet to decode is clearly not a simple linear list of instructions, but a program, with subroutines, callbacks, loops, and all the complexity that implies, so that one can talk of the further possible combinations of instructions — the additional combinatorial complexity — of our genome over the nematode's. Moreover the genes represent only a small part of the genome, and the rest, including what has been injudiciously termed junk DNA, probably contains a great deal of information (following our computing analogy, this may include variable and constant definitions, and so on). The second is that the genome only specifies what it must to get the job done, and so, if without further genetic intervention, a certain process will take place following the evolution of a physical or chemical system alone, then nature may take advantage of this; the genes act just as the choreographers. Some processes are more tightly choreographed than others: we now know that segmentation in *Drosophila* is a process tightly genetically controlled by a number of genes, yet other similar processes have much looser control³ We need to understand the theoretical basis of biological complexity.

2. Beyond genomics

The discovery of the double helix of DNA determined a clear link between genetics and coding theory, as Crick and Watson realized immediately: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material" [12]. The biochemical puzzle of the genetic code was thus reduced to an abstract problem of symbol manipulation, as other physical scientists also saw, and for a few years physical scientists made large contributions to resolving the mysteries of genetics⁴ However, after the decipherment of the standard genetic code in the 1960s, the attention of most DNA research gradually turned to the experimental aspects related to the sequencing of genomes of different organisms. These efforts led to the sequencing of the whole human genome; the Human Genome Project triggered strong hopes, inter alia, of the possibility of diagnosing and treating many serious diseases. However, more than a decade after its completion in 2003, it is generally acknowledged that these expectations have not been met. A 2010 article by S.S. Hall that appeared in *Scientific American*, "Revolution Postponed" asked why the human genome project had been disappointing" [17]:

The problem is that research springing from the genome project has failed as yet to deliver on the medical promises that Collins and others made a decade ago. Tumor biologist Robert A. Weinberg of the Whitehead Institute for Biomedical Research in Cambridge, Mass., says the returns on cancer genomics "have been relatively modest — very modest compared to the resources invested." Harold E. Varmus, former director of the National Institutes of Health, wrote recently in the *New England Journal of Medicine* that "only a handful of major changes ... have entered routine medical practice" — most of them, he added, the result of "discoveries that preceded the unveiling of the human genome." Says David B. Goldstein, director of the Center for Human Genome Variation at Duke University: "It's fair to say that we're not going to be personalizing the treatment of common diseases next year."

What are the possible causes of this disappointment? On the one hand, many scientists think that we do not yet have enough information available; that more data are required. This line of thought has led to the 1000 Genome Project. While this "big data" paradigm has undeniably led

³This point was argued in a 2012 issue of *Philosophical Transactions* 'Beyond crystals: the dialectic of materials and information' [11].

⁴As the works of Rumer newly translated in this volume attest [13–16]. The RNA tie club, cofounded by Watson and George Gamow, with members such as Francis Crick, Richard Feynman, and Max Delbrück trained in the physical sciences, played with these interdisciplinary ideas in the early years of DNA research.

to important technological advances and has been also a source of inspiration for methodological improvements, some researchers, including ourselves, think that, in parallel, we need major advances in the theoretical interpretation of such data, including a more holistic approach to biology. The complexity of many biological problems makes black box computational approaches useless without theoretical guidance. Hence, we argue that in addition to the “big data” enterprise, we need a new paradigm necessarily grounded on first principles from mathematics, physics, and chemistry.

3. The mathematization of biology

Biology at present is embarked on an experimental search that we may define as functionalist; that is to say that it is attempting to understand how the functions of living material link together. This search is based upon data that are harder and harder to classify, and above all to interpret. We may compare the situation to that of the comprehension of inanimate matter before the advent of the modern atomic theory. We may thus ask ourselves: were those theoretical efforts to understand and classify matter using physico-mathematical concepts useful? The answer is of course affirmative, and indeed theoretical methods used by biology today originated in the revolution — the paradigm shift — produced by the knowledge of the atomic structure of matter, without which molecular biology would not exist. We argue that another paradigm shift is needed to understand biology: its mathematization. It is held by some biologists today that mathematics will never be useful in their discipline in the way it has been in other sciences. But to argue this is to misunderstand the nature of mathematics. Mathematics at its core is no more than the search for patterns in structures; as all is fundamentally patterns, mathematics is universal. In the same way that other sciences have mathematized themselves, biology must surely follow; as Newton did for physics, Dalton for chemistry and Fisher for statistics so it is necessary to undertake this project for biology. In other words, biology needs its own particular Newtonian revolution. For this it is imperative that the vast accumulation of data coming out of biological research should begin to be organized into a logical structure. But for this end it is not generally useful to attempt to transfer *manu militari* the methods of mathematics that have served physics. Within the present secrets of life there probably lies much more mathematics than has yet been discovered. Thus the paradigm shift is like the end of the rainbow, seemingly remaining out of reach even as we move towards it, and it may be for this reason that some have viewed biology as bound to remain essentially inaccessible to the methods of reasoning of physics and mathematics.

We however, are convinced that this is not the case: the success of science in explaining the universe over the past few centuries has come by travelling rails composed of the parallel paths of experiment and theory. Biology is to a great extent lacking this theory, while the experimental rail has pushed far ahead. To carry the analogy further, the presence of sleepers or ties between the two rails, that is to say interrelations between systematic rational thought on one hand and experimental observations on the other, has been fundamental to the success of other disciplines of science. In this sense we may compare biology to technology, which can and does also move forward without science, as it is possible to develop a technology without understanding its conceptual basis. But experimental biology without a conceptual basis, just like blind technology, is as dangerous as a monkey with a pistol! The experimental aspects need a strong quantitative theoretical background and we should learn from history.

For example, the modern theory of dynamical systems studies how phenomena such as chaos, self-organization, synchronization, complexity, and patterns, can emerge from nonlinear interactions within a system. This field comprises a growing set of powerful tools with which nonlinear phenomena in any field of natural science may be analysed, and as it looks for universal behaviour of a system it provides a systematic way of approaching a problem that can unite disparate observations. The procedure is both reductionistic and at the same time holistic; first one must strip a problem down to its essence, and then one can see how interactions between the parts lead to the emergence of complex behaviour of the whole that is not present in any one part. The nonlinear revolution is still taking place; at present nonlinear dynamics has only

partially been assimilated into the edifice of science, as the foundations of the field are not yet complete, and there is still much to understand regarding the fundamentals of many aspects of nonlinear systems. But it is from these ideas that we see the, as yet nonexistent, theoretical biology germinating. The revolution or paradigm shift these ideas are producing in science is comparable to that when when statistical methods were introduced⁵. From this revolution there is coming another set of multipurpose implements that, like the earlier statistical ones, can be applied to problems in all areas of science. With them can be understood phenomena that arise from nonlinear interactions between the elements of a system. Since nonlinear systems are the norm, and linear systems — the more studied in the past — are rarities in science, the analysis of nonlinear dynamical systems leading to complex behaviour is encountering a vast ocean of applications throughout the whole of science. (Studying linear systems is like studying the biology of a particular animal, say an elephant. On the other hand studying nonlinear systems is like studying all other biological systems that aren't elephants. Just as it doesn't make much sense to organize biological studies into elephant and non-elephant biology, neither is the label *nonlinear* dynamics anything more than a reminder that in the past we have placed too much emphasis on the very particular mores of linear systems.) Despite this vast scope, it is becoming clear that a small number of organizing rules govern the behaviour of disparate systems in apparently unrelated fields of science. The mathematical and physical principles behind this tendency to pattern formation or self-organization in nature are what dynamical systems can tell us about.

We insist that the theoretical rail of biology — of dynamical systems, of other mathematics, and of mathematics not yet conceived — needs to be constructed, and that what will be necessary to that end is to carry forward concurrently the programmes of mathematizing biology and of developing the necessary mathematics. Roger Bacon, writing in 1267, put the point beautifully: “Et harum scientiarum porta et clavis est mathematica” — mathematics is the door and key to the sciences [19], and as a statistician, Karl Pearson, said in 1901 [20] “I believe the day must come when the biologist will — without being a mathematician — not hesitate to use mathematical analysis when he requires it.”

4. The nature of biological information

At least a part of this new paradigm must be based on understanding biological information, where it is found, and how it evolved. One month after their first paper on DNA, Crick and Watson wrote in their second publication that “it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information” [21]. Shortly after that George Gamow wrote [22]

In a communication in *Nature* of May 30, p. 964, J. D. Watson and F. H. C. Crick showed that the molecule of deoxyribonucleic acid, which can be considered as a chromosome fibre, consists of two parallel chains formed by only four different kinds of nucleotides. These are either (1) adenine, or (2) thymine, or (3) guanine, or (4) cytosine with sugar and phosphate molecules attached to them. Thus the hereditary properties of any given organism could be characterized by a long number written in a four-digital system. On the other hand, the enzymes (proteins), the composition of which must be completely determined by the deoxyribonucleic acid molecule, are long peptide chains formed by about twenty different kinds of amino-acids, and can be considered as long ‘words’ based on a 20-letter alphabet. Thus the question arises about the way in which four-digital numbers can be translated into such ‘words’.

How can biological functions be included in the framework of information theory? — some people refer to this point as introducing meaning into information theory. Classical

⁵Ronald Fisher, the father of modern statistics, introduced the majority of the techniques that are now commonly accepted as standard when facing problems related to genetics, biology and agriculture. He is also considered one of the three founders of population genetics. In 1938 he quipped [18]: “To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.”

information theory disregards the content that the physical information carries; information and communication theory deal only with secure transmission of information through noisy channels. However, if we consider the genetic code we can observe some peculiarities that deviate from the classical paradigm: i) the message to be sent and the sender are intimately related; ii) information includes how to produce new senders and receivers; and iii) the messages are not arbitrary, that is, encoding, transmitting, and decoding mechanisms should be best fitted for the finite (albeit large) set of messages to be transmitted.

A common metaphor refers to DNA as the “book of life”. Of course, we know that the main information that represents an organism is contained or carried by nucleic acid molecules. In this respect, DNA can be considered as a book, but curiously, such a metaphor has a scientific basis only in the concept of the genetic code. However, the genetic code is not a book nor a part of it; rather it is a translation dictionary between two different worlds (languages), i.e., the world of nucleic acids and the world of proteins. Hence, the genetic code allows the translation of a book written in a language into an abridged version of the same book in a different language. Moreover, little is known about the grammar, the syntax, and even the orthography of the book of life. Still, we know that the genetic code is involved in the transmission of the information contained in such book and configures a relevant part of the process that defines the central dogma of molecular biology. The term code, as it is used in biology, arises from information and communication theory. From the birth of communication theory, codes have been used with a twofold scope, i.e., to translate some information for transmission, for example as in radio communication, and to allow error detection and correction. It must be remarked that, in any kind of communication system, errors are unavoidably introduced in the transmission process. To date, our knowledge of the genetic code concerns only the first aspect, the error correction features being interpreted only in terms of robustness to mutations without accounting for the informational aspects.

However, the most recent insights in the formal aspects of the genetic code show that there is a high degree of mathematical structure, in the form of arithmetical and algebraic properties. This prompts compelling questions about the genetic code: does it carry additional information other than that related to specific coding of the chain of amino acids that form a protein? Is it a fundamental part of a complex decoding system that uses algebraic and group-theoretic properties for detecting and correcting informational errors? These apparently simple questions pose a difficult problem that requires a strong multi-disciplinary effort. Still, they represent only the most elemental aspects related to the flux of biological information. In analogy with the book of life, they are related to the orthographic level. A step beyond is represented by syntax; how are words concatenated into sentences and paragraphs? A crucial aspect of this problem has been pointed out by the discovery of (trinucleotide) circular codes in genes of prokaryotes and eukaryotes. Circular codes allow the retrieval of the frame of construction (or reading frame) in any position of any sequence generated by them. They possess a strong algebraic properties and are related to the necessity of maintaining the correct reading frame of coding sequences in protein synthesis. However, the elemental mechanisms that take advantage of such structures are unknown.

At the heart of the matter of DNA and information lie questions of this type. Some of the specific developments that we believe hold promise to solve these open questions are discussed in the papers making up the theme issue. In this theme issue are published a series of papers giving the state of the art surrounding DNA and information. The issue has a strong interdisciplinary character, bringing together scientists from different fields such as biology, mathematics, statistics, physics, and engineering. The aim of the editors is that the issue should stimulate discussions and scientific exchange that will lead to new research on why and how life can exist from the point of view of the coding and decoding of genetic information. The issue contains also two opinion papers that review some of the contributions: the piece by Philip Ball [23] focuses on the concept of information in biology and include comments on the papers by [24–31]; the piece by Peter Wills [32] discusses the matter of information in biology also in an evolutionary context by taking into account the problem of the origin of life and of the genetic code. The contributions

commented are those of [25–31,33–37]. Ideally, the organization of the papers in the issue follows that of the above reviews which we placed after them. The issue closes with the three seminal articles by Yury Borisovich Rumer on the symmetry of the genetic code that have been expressly translated from Russian for this publication [13–16].

References

1. Herschel JFW. 1830 *A Preliminary Discourse on the Study of Natural Philosophy*. Longman, Rees, Orme, Brown and Green, London.
2. Visconti GA, Visconti EQ. 1788 *Il Museo Pio-Clementino*, volume 4, p. plate 13. Luigi and Giuseppe Mirri, Rome.
3. Crick F. 1958 On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163.
4. Crick F. 1970 Central dogma of molecular biology. *Nature* **227**, 561–563.
5. Crick F. 1988 *What Mad Pursuit: A Personal View of Scientific Discovery*. Basic books.
6. Stewart P. 1964. *Jacobellis v. Ohio*, 378 U.S. 184.
7. Brenner S. 1991 Summary and concluding remarks. In *Evolution of Life* (ed. S Osawa, T Honjo). Springer.
8. Claverie JM. 2001 What if there are only 30,000 human genes? *Science* **291**, 1255–1257.
9. C elegans Sequencing Consortium. 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018.
10. Pellicer J, Fay MF, Leitch IJ. 2010 The largest eukaryotic genome of them all? *Botan. J. Linnean Soc.* **164**, 10–15.
11. Cartwright JHE, Mackay AL. 2012 Beyond crystals: the dialectic of materials and information. *Phil. Trans. R. Soc. A* **370**, 2807–2822.
12. Watson JD, Crick FHC. 1953 A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.
13. Fimmel E, Strüngmann L. 2016 Yury Borisovich Rumer and his ‘biological papers’ on the genetic code. *Phil. Trans. Roy. Soc. A* **this issue**.
14. Rumer YB. 2016 Systematization of codons in the genetic code (a). *Phil. Trans. Roy. Soc. A* **this issue**.
15. Rumer YB. 2016 Systematization of codons in the genetic code (b). *Phil. Trans. Roy. Soc. A* **this issue**.
16. Rumer YB. 2016 Systematization of codons in the genetic code (c). *Phil. Trans. Roy. Soc. A* **this issue**.
17. Hall SS. 2010 Revolution postponed. *Sci. Am.* **October**, 60–67.
18. Fisher RA. 1938 Presidential address, first Indian statistical congress. *Sankhyā: The Indian Journal of Statistics* **4**, 14–17.
19. Bacon R. 1267 *Opus Majus*.
20. Sutherland J. 1901 Education in science. *Nature* **63**, 275–276.
21. Watson JD, Crick FHC. 1953 Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**.
22. Gamow G. 1954 Possible relation between deoxyribonucleic acid and protein structures. *Nature* **173**, 318.
23. Ball P. 2016 The problems of biological information. *Phil. Trans. Roy. Soc. A* **this issue**.
24. Barbieri M. 2016 A new theory of development. *Phil. Trans. Roy. Soc. A* **this issue**.
25. Barbieri M. 2016 What is information? *Phil. Trans. Roy. Soc. A* **this issue**.

26. Fimmel E, Michel C, Strüingmann L. 2016 n -nucleotide circular codes in graph theory.
Phil. Trans. Roy. Soc. A **this issue**.
27. Koonin E. 2016 The meaning of biological information.
Phil. Trans. Roy. Soc. A **this issue**.
28. Roederer J. 2016 Pragmatic information in biology and physics.
Phil. Trans. Roy. Soc. A **this issue**.
29. Tlusty T. 2016 The self-referencing DNA: a remark on physical and geometrical aspects.
Phil. Trans. Roy. Soc. A **this issue**.
30. Varn D, Crutchfield J. 2016 What did erwin mean? the physics of information from the materials genomics of aperiodic crystals & water to molecular information catalysts & life.
Phil. Trans. Roy. Soc. A **this issue**.
31. Walker S, Kim H, Davies P. 2016 The informational architecture of the cell.
Phil. Trans. Roy. Soc. A **this issue**.
32. Wills P. 2016 DNA as information.
Phil. Trans. Roy. Soc. A **this issue**.
33. Adami C. 2016 What is information?
Phil. Trans. Roy. Soc. A **this issue**.
34. Bartholomäus A, Fedyunin I, Feist P, Sin C, Zhang G, Valleriani A, Ignatova Z. 2016 Bacteria differently regulate mRNA abundance to specifically respond to various stresses.
Phil. Trans. Roy. Soc. A **this issue**.
35. Gonzalez D, Giannerini S, Rosa R. 2016 The non-power model of the genetic code: a paradigm for interpreting genomic information.
Phil. Trans. Roy. Soc. A **this issue**.
36. Paci G, Cristadoro G, Monti B, Lenci M, Esposti MD, Castellani G, Remondini D. 2016 Characterization of DNA methylation as a function of biological complexity via dinucleotide inter-distances.
Phil. Trans. Roy. Soc. A **this issue**.
37. Wills P. 2016 The generation of meaningful information in molecular systems.
Phil. Trans. Roy. Soc. A **this issue**.