BioEssays
www.bioessays-journal.com

# Replicating and Cycling Stores of Information Perpetuate Life

*Antony M. Jose*

**Life is perpetuated through a single-cell bottleneck between generations in many organisms. Here, I highlight that this cell holds information in two distinct stores: in the linear DNA sequence that is replicated during cell divisions, and in the three-dimensional arrangement of molecules that can change during development but is recreated at the start of each generation. These two interdependent stores of information – one replicating with each cell division and the other cycling with a period of one generation – coevolve while perpetuating an organism. Unlike the genome sequence, the arrangement of molecules, including DNA, RNAs, proteins, sugars, lipids, etc., is not well understood. Because this arrangement and the genome sequence are transmitted together from one generation to the next, analysis of both is necessary to understand evolution and origins of inherited diseases. Recent developments suggest that tools are in place to examine how all the information to build an organism is encoded within a single cell, and how this cell code is reproduced in every generation.**

## 1. Introduction

One of the amazing aspects of living things is that they transmit the information for building themselves from one generation to the next. While much of what an organism is made of depends on the linear sequence information in its DNA genome, this sequence is not the only store of information that is transmitted across generations. We can see evidence for the transmission of extra-genomic information in cases where changes that do not alter DNA sequence nevertheless persist for many generations (see **Box 1** for a classic example). Such inheritance of extra-genomic changes invites a consideration of the scope and formulation of all the inherited information that specifies the developing organism's ensemble of traits – not only the information encoded in the sequence of bases in the DNA, but also the non-genetic information encoded in molecular assemblies independent of DNA sequence.

The information needed to perpetuate an organism must be present even in the life stage that has the least number of cells. This stage serves as a bottleneck for the transmission of information from one generation to the next and is minimally a single cell. Therefore, all the molecules and their arrangement that is reproduced in every such bottleneck stage of an organism is the minimal information that specifies the perpetuation of that organism. This minimal information includes the genome sequence and a spatial arrangement of the genome along with other pre-existing molecules, including RNAs, proteins, sugars, lipids, etc. Organisms transmit these two stores of information from one generation to the next using different strategies. The linear DNA sequence is transmitted by replicating it at each cell division. This strategy allows for retention throughout development of largely the same sequence in every cell that forms a continuous lineage from one generation to the next. The three-dimensional arrangement, on the other hand, is transmitted by cycling it with a period of one generation. This strategy allows for potentially extensive transformation of the arrangement throughout development before returning to a similar configuration in the next generation. In addition to changes in the arrangement of molecules, components could be replenished or added to in each generation using the DNA template and raw materials from the environment over the course of a life cycle. Thus, life is perpetuated by two stores of information that are interdependent yet distinct.

The need for analyzing both forms of information together is evident if we consider a simple regulatory loop where a transcription factor binds its own promoter and sustains its production in every cell. At the start of each generation, in addition to general cellular machinery, this "developmental program" needs the DNA sequence encoding for the transcription factor, the promoter sequence that the transcription factor binds to, *and* the transcription factor itself arranged such that the transcription factor can bind to its binding site in the promoter. In other words, developmental programs are specified using DNA sequence and an arrangement of additional pre-existing molecules (see Refs. [1–6] for related views).

Here, I develop this perspective on inherited information further, discuss its implications, and suggest approaches for the discovery of the stores of information that are transmitted along with the genome sequence across generations to drive development and evolution (see **Box 2** for key concepts). While this article has an emphasis on animal biology to give it focus, the concepts presented here are applicable to all cellular life: bacteria, archaea, and eukarya.

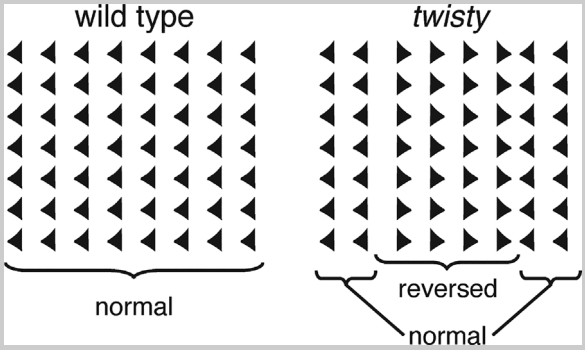
Prof. A. M. Jose
Department of Cell Biology and Molecular Genetics
University of Maryland
College Park, MD 20742, USA
E-mail: amjose@umd.edu

**DOI: 10.1002/bies.201700161**

## Box 1. Cortical inheritance

Classic experiments illustrating the inheritance of changes in non-genomic information were performed using the ciliate *Paramecium aurelia*.[104] These organisms have hair-like cilia in their cell cortex that help them swim and that arise from cortical units arranged in regularly spaced rows (Figure, Box 1). The first hints that changes in these cortical patterns could be inherited came from the analysis of a spontaneous swimming mutant called *twisty*. Unlike normal cells, *twisty* cells had a patch of four rows with reversed polarity of cortical units, flanked by altered spacing – a defect that could be observed even after ≈300 generations (Figure, Box 1). This defect was deduced to be the result of a patch of cortex acquired during sexual reproduction by one cell of a conjugant pair from the other cell oriented at 180°. Armed with an explanation for how *twisty* was generated, various grafts of a piece of one cell on a whole cell were generated by letting similar conjugating pairs separate spontaneously such that a part of one conjugant went with the other or by removing most of one conjugant with a micromanipulator, leaving a residual graft on the other. The resultant cells that had rows of cortical units with reversed polarity could be maintained indefinitely with periodic selection. Even those that reverted to normal cortical structure had transmitted the variant organization to a remarkably large number of progeny cells ($10^9$–$10^{12}$). These observations illustrate the principle that pre-existing structure can inform the newly generated structure in a cell (see Ref. [5] for additional examples).




Cortical architecture of wild-type and *twisty* cells. The cortical units from which cilia arise can be seen using silver nitrate staining as regularly spaced units with left-right asymmetry (point vs. base of each triangle). Left) schematic of a patch of cortex from wild-type cells. Right) schematic of a patch of cortex from *twisty* cells showing four rows with reversed polarity flanked by rows with normal polarity (based on Ref. [104]).

## 2. DNA Proposes, Cell Disposes

While the genome is a repository of all sequences that can be transcribed and used for making other components, at any given moment, which RNA is transcribed from DNA depends on what else is within the cell and on the cell's interactions with the external environment. The information contained in the genome of an organism is thus not sufficient to make that organism. To appreciate this insufficiency, consider a single cell in an organism: (a) the DNA within this cell does not encode all aspects of all molecules in the cell and (b) whether a molecule is made using the DNA depends on other contents in the cell.

### 2.1. The Linear DNA Sequence is Not Sufficient to Build the Three-Dimensional Cell

The RNA and protein components of a cell ultimately rely on DNA for production through the process of transcription and translation using smaller molecules, nucleotides, and amino acids, as raw materials. These proteins and RNA can go on to catalyze the formation of additional components such as lipids and sugars from other raw materials. However, the three-dimensional structure and spatial arrangement of all the molecules within a cell is not encoded in the linear DNA. For example, where a protein is localized in the cell can change over time based on interactions with other components within a cell.[7] While the DNA encodes the sequence of a protein, the structure it takes depends on its immediate environment.[8] Environmental changes can even induce cells to over-ride defects caused by mutations in their DNA. For example, mutation of the GTPase Rac1 can invert the polarity of epithelia and yet normal polarity can be restored by the addition of exogenous laminin protein.[9] Components in organelles such as mitochondria are not entirely encoded by the nuclear genome and can be considered as endosymbionts with partial autonomy within a eukaryotic cell.[10,11] Finally, the sizes, numbers, and shapes of organelles, and indeed the collective architecture of a cell all depend on dynamic interactions between components within the cell and with its external environment.[12]

### 2.2. Many Cell Types Use the Same Genome

The presence of different cell types that retain their identity over time and across cell divisions within an organism reveals that the contents of a cell can exert a controlling effect on the DNA. The mechanisms that make and maintain different cell types were initially referred to as epigenetic control systems[13] to signify that they are above ("epi") genetic control. Pioneers studying such control mechanisms in bacteria imagined multiple modes of regulation for components within a cell in the context of biochemical reactions,[14] many examples of which have been characterized in the past half-century. These regulatory modes provide programmed constraints that explain "how come?" when events happen within a cell.[15] Furthermore, computational exploration suggests that even random networks of elements with high molecular specificity can result in the emergence of different cell states that remain stable over time.[16]

## Box 2. Key Concepts

*Replicating stores of information*: These are molecules with measurable features that are propagated through templated replication. Genomic DNA with its sequence information is the best characterized example of such a store. However, other non-genetic information can also be propagated through templated replication. For example, the inheritance of alternative folding states of proteins like prions illustrates the transmission of structural information through replication, independent of DNA sequence.[51]

*Cycling stores of information*: These are molecules with measurable features that cycle such that their quantity, quality, and/or localization can change over time but return periodically to a similar configuration. The concentration and overall three-dimensional arrangement of DNA, RNAs, proteins, ions, metabolites, chemical modifications, and potentially much more constitute cycling stores of information. Importantly, the information contained in the collective arrangement of molecules at any given time is a function of the DNA sequence and the collective arrangement at an earlier time. The extent of heritable variation depends on the fidelity with which cycling stores of information are recreated in every generation.

*Transgenerational homeostasis*: This is the process that opposes inheritance of non-genetic changes and preserves the ancestral configuration. Non-genetic changes that occur in one generation, especially within the germline, can sometimes persist in subsequent generations – a phenomenon referred to as transgenerational epigenetic inheritance.[55] However, these changes are often lost after a few generations,[59–65,67–77] ensuring preservation of the ancestral configuration (see Ref. [105] for a discussion of homeostasis and evolution).

Extensions of these ideas have led to the identification of gene regulatory networks in model organisms,[17] where collections of active genes and gene products maintain different cell types while being able to respond to signals during development.

## 3. Replicating and Cycling Stores of Information Together Drive Development

Since its beginnings as *Entwickelungsmechanik* more than a century ago,[18] the rich field of developmental biology has been addressing various aspects of how an organism is made. But, the minimal information that is necessary to specify the development of a particular organism in a given external environment is still unknown, and is in fact, relatively unexplored.

### 3.1. An Integration of Past Events Influences the Future

Consider an organism that progresses from one generation to the next through a single-cell stage, as is the case with humans and most known multicellular organisms. To discover the minimal information that is necessary for the development of this organism, the entire life cycle of the organism needs to be examined (**Figure 1**A). For chicken, this would mean examining from egg-to-egg and not from egg-to-hen (developmental biology) or from hen-to-egg (reproductive biology). The molecules that could store information within a cell are expected to have accrued over time based upon the action of pre-existing molecules on DNA in a changing extracellular environment (Figure 1B, top). This process of iterative accumulation of material and progressive change in the state of a cell implies that the DNA *and* everything else accrued in the past that is now within the cell together predicts the next state of a cell in a given environment. Therefore, the single cell that starts each generation must have a similar three-dimensional arrangement of molecules to ensure the development of similar organisms in successive generations (Figure 1B, bottom). This minimal information that encodes the making of an organism and is contained within a single cell can be considered as the *cell code* for the organism (see **Box 3**).

Early attempts at unifying biology focused on understanding the cell in development and inheritance.[19] More than a century ago, some experimental embryologists clearly appreciated the iterative progression (e.g., Ref. [20]) and influence of history (e.g., Ref. [21]) when describing the nature of organisms. The interdependence of genes and non-genetic factors within a cell was also well appreciated (e.g., Refs. [22,23]). In fact, the term "epigenetics" was initially defined as ". . . causal mechanisms by which the genes of the genotype bring about phenotypic effects,"[24] but later stated as "the causal interactions between genes and their products which bring the phenotype into being,"[25] making it clear that the genotype (DNA) alone does not bring the phenotype into being. These insights from embryology were also recognized by early proponents of the Modern synthesis that combined Mendelian genetics with Darwinian evolution,[26] and when ignored, can lead to the mistaken popular view that DNA is the blueprint of life.

In the context of an organism, the contents of a cell can change dramatically over time independent of changes driven by the DNA within the cell. For example, consider progression from one generation to the next in the worm *Caenorhabditis elegans*[27] (see Box 3). From the single-celled zygote, two primordial germ cells are established five cell divisions later. These two cells then go through an extended period of quiescence while the rest of the organism develops. During this period intestinal cells cannibalize a large portion of the cytoplasmic contents of both cells.[28] Then, the cells proliferate to generate germ cells that subsequently differentiate, first producing sperm, and then oocytes. Maturing oocytes acquire cytoplasmic contents from the rest of the germline[29] and yolk from the intestine,[30] potentially along with extracellular RNAs.[31–33] These acquisitions make the oocyte larger than all the 558 cells of a hatching larva combined. Fertilization of this enlarged oocyte creates the zygote for the next generation. Both

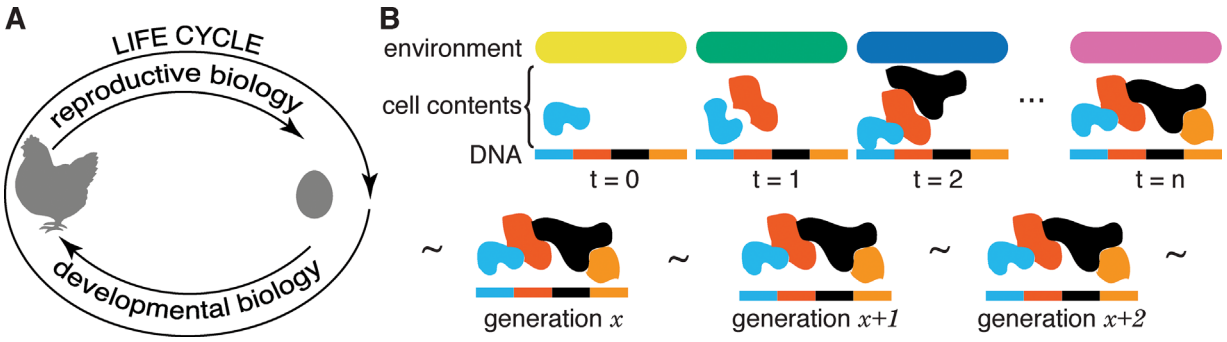



**Figure 1.** Perpetuation of life. A) Understanding what perpetuates life requires examination of entire life cycles in addition to the study of how organisms are made (developmental biology) and how they reproduce (reproductive biology). B) Accrual of molecules in living systems. *Top*, Contents of a cell accrue through sequential production of different gene products and the molecules they modify (colored shapes matching color in DNA) while the environment of the cell (rounded bar) and previously made components continue to change over time. *Bottom*, Life is perpetuated through the recreation of a similar arrangement of cellular contents at the start of each generation.

dramatic changes in this cycle – loss of cytoplasmic material from primordial germ cells and gain of cytoplasmic material by oocytes – occur during periods of relative transcriptional quiescence.

### 3.2. What is the Minimal Information Necessary to Make and Perpetuate an Organism?

The upper limit for the cell code is the entire contents and their arrangement in the zygote that begins each generation. However, the cell code need not equal everything in a zygote. If only a subset of the components and arrangement within a zygote are reproduced in every generation (**Figure 2**), then the cell code can be less than the contents of a zygote. Furthermore, because many molecular assemblies in life are capable of self-organization (e.g., mitotic spindle[34]) and templated processes (e.g., transcription[35]), it is possible that essentially the same cell code could be specified using a subset of the molecules and arrangement that are recreated in every zygote. This possibility is supported by experiments on single-cell regeneration in the ciliate *Stentor polymorphus* – 1/27th of its initial volume can regenerate all structural features.[36]

Determining the minimal representation of the cell code of an organism requires minimization of features in a zygote akin to current efforts to minimize the genomes of cells.[37,38] For example, imagine a gene that has two copies of partially transcribed RNA, 32 copies of the spliced RNA, 17 copies of translated protein, 12 copies of the protein with correct subcellular localization, etc., that are all reproduced in the zygote of every generation, i.e., that are all part of the cell code. Which of these details are necessary for the perpetuation of a similar arrangement? What other arrangements of these components are essentially equivalent? What changes to the rest of the zygote can impact this gene and the arrangement of its products while perpetuating life?

A problem that is separate from determining the cell code of an organism is determining how it is reproduced in every generation. Reproduction of the genetic component of the cell code, that is, DNA sequence, is relatively well-understood and is through the process of templated replication at each cell division. Reproduction of the non-genetic components of the cell code, that is, the spatial arrangement of molecules including DNA, is likely to be more complex. One possibility is that the essential arrangements are preserved through cell divisions despite differentiation and development of the organism such that a continuous line of cells from the zygote of one generation to the next carries the cell code for the organism. Alternatively, and more likely, these arrangements vary during differentiation and development but go through an elaborate cycle such that they return to a similar configuration in the zygote of each generation. In the case of unicellular organisms, the cell code needs to be reproduced at a certain time in each cell division cycle. In organisms such as plants where many cells are capable of generating the entire organism, every such cell needs to be able to reproduce the cell code when exposed to conditions that stimulate differentiation and development. Currently, we have a fragmentary understanding of changes in a few potential components of the cell code during development and are far from understanding the mechanism(s) by which the cell code is reproduced in any organism. Nevertheless, a range of components and arrangements are likely to be nearly equivalent cell codes for an organism. We already know that the genome sequences of individuals within a species can vary a little and yet result in the generation of similar organisms. Additionally, discovering alterations to non-genetic aspects of the cell code that are compatible with the perpetuation of life will reveal the full extent of novelty that can arise from one generation to the next in an organism.

## 4. Persistent Ancestral Information Modifies the Cell Code

A key concern in thinking about how organisms evolve is what changes in one generation can be passed on to subsequent generations. The nature of organisms presented here (Figure 1 and 2) makes it clear that the persistence of changes across many generations requires modification of the cell code and thus evolution occurs through "descent with modification"[39] of the cell code. Note that this does not include information (induction

## Box 3. Cell code

Many multicellular organisms cycle through a single-cell bottleneck – the fertilized zygote in sexually reproducing organisms – that carries all the information necessary to make the next generation.[19] In animals that lay eggs, all the information necessary for development is contained within the egg. In animals that give birth to live young, the fertilized zygote grows into a fetus that shares circulation with the mother. In this case, the growing fetus could potentially acquire additional information from the mother after the single-cell bottleneck. Organisms that undergo parthenogenesis[106] also cycle through a single cell, an egg that does not need fertilization by sperm. In contrast, parasites go through elaborate cycles within different cells and even different host organisms,[107] and regenerating organisms can begin each generation from many different collections of cells.[108] Nevertheless, in every scenario, the simplest life stage of an organism *minimally* consists of a single cell that encodes all the information required for making the organism in the next generation.

To clarify the relationship between the contents of a single cell and its genome *sequence*, let us define all molecules (including the DNA genome) and their spatial arrangement in a single cell as "C" and environmental inputs as "E" (Figure Box 3A). These external inputs include factors that influence the cell (e.g., temperature), interact with the cell (e.g., extracellular signals), or are imported into the cell (e.g., nutrients). At an arbitrary initial time, $t = 0$, $C_0$ indicates the state of molecules present in the cell, $E_0$ indicates the external inputs, and • indicates their interaction. At a later time, $t = 1$, $C_0$ could have changed ($C_{0\rightarrow1}$) and can now interact with, modify, and/or be modified by the new $C_1$ made at $t = 1$ ($C_{0\rightarrow1}\bullet C_1$). Similar changes could also occur in the imported or interacting external material ($E_{0\rightarrow1}\bullet E_1$). Furthermore, both the accrued cellular components ($C_{0\rightarrow1}\bullet C_1$) and the changing external input ($E_{0\rightarrow1}\bullet E_1$) can interact and/or modify each other ($C_{0\rightarrow1}\bullet C_1\bullet E_{0\rightarrow1}\bullet E_1$). Thus, the changing contents and their arrangement in a cell are not predictable from the sequence of its genome alone. This influence of the past on the current state of the cellular contents and their arrangement implies that a cell at time $t = n$ includes information stored in its genome sequence ($DNA_{seq}$) and in the arrangement of contents accrued since the origin of cellular life ($C_{past\rightarrow n}\bullet E_{past\rightarrow n}$). If such a single cell is a zygote that begins each generation, then everything in this cell ($DNA_{seq} + C_{past\rightarrow n}\bullet E_{past\rightarrow n}$) that is nearly reproduced in the zygote of successive generations is the upper limit of the cell code for the organism (Figure Box 3B). This recreation of similar cell codes in successive generations perpetuates life and is a hallmark of organisms that constrains their evolution.

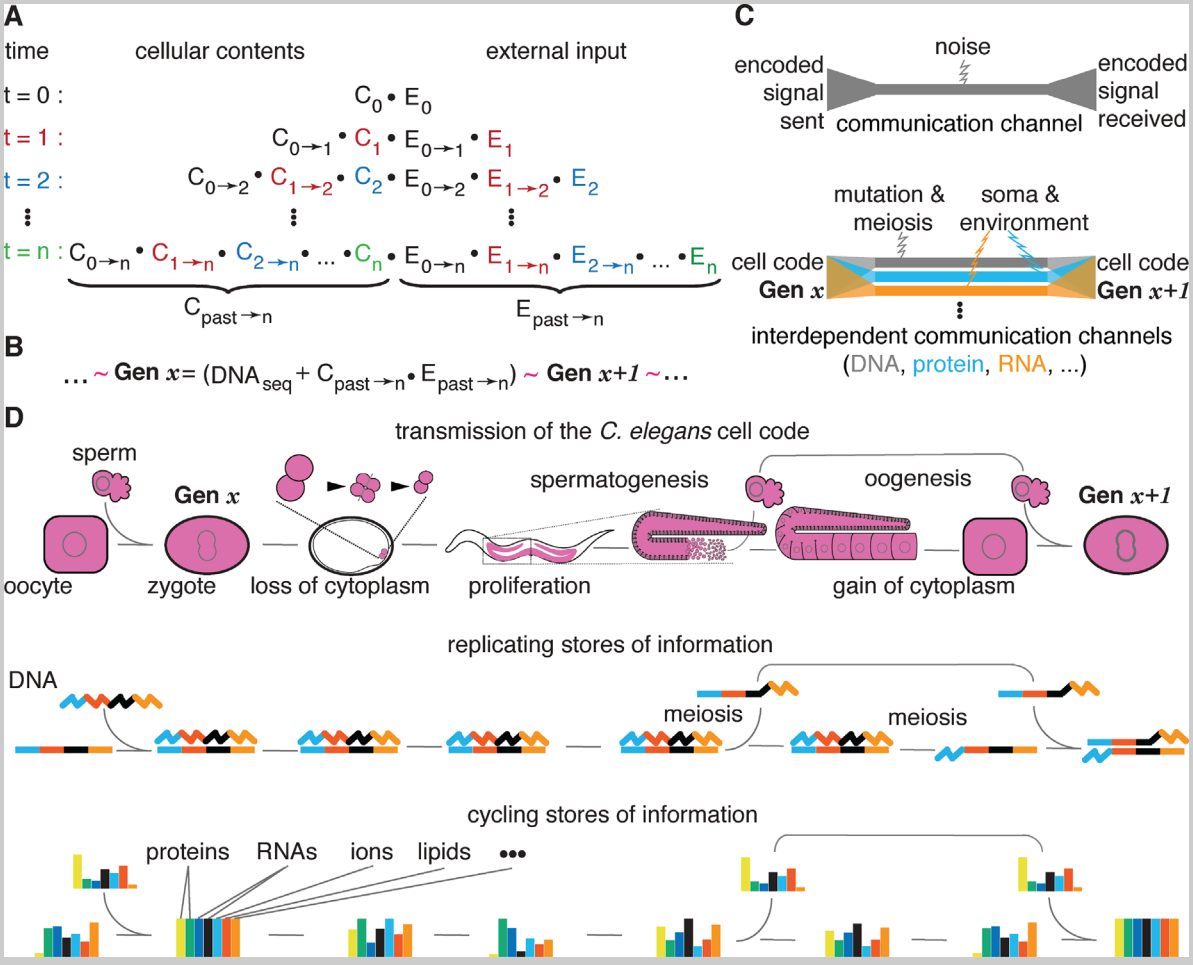
Transmission of the cell code from the zygote of one generation to the zygote of the next generation through a developing organism is akin to transmission of coded signals from a sender to a receiver through a communication channel (Figure Box 3C), which has been given mathematical form in information theory.[109] In this comparison, the pathways through which individual aspects of the cell code (DNA, RNAs, proteins, spatial arrangements, etc.) are transmitted from one generation to the next are analogous to communication channels, whereas, systematic and stochastic external influences are analogous to noise. However, application of information theory to study cell codes additionally requires accounting for the following characteristics:

*1. Interdependence of channels*. The channels through which the cell code is transmitted from one generation to the next are related to each other through complex functions. For example, consider transmission of the information specifying the three-dimensional structure of a protein. The primary sequence of this protein depends on how the transcribed RNA is spliced into mRNA. The secondary and tertiary structures of the protein depend on the composition and abundance of molecules in its immediate vicinity (ions, sugars, etc.).

*2. Dynamics of information in transit*. Cycling stores of information also include unstable molecules and potentially ephemeral interactions that need to be transmitted from one generation to the next. Therefore, the encoded information must exist either in multiple interconvertible configurations or be transferrable to other molecules during development.

Systematic and stochastic changes can occur during the transmission of a cell code from one generation to the next. The nematode *C. elegans* provides a well-characterized example of such "noise" during transmission (Figure Box 3D). The DNA sequence is subject to stochastic exchange between homologs during the two meioses that generate gametes (sperm and oocyte). The cytoplasmic contents undergo dramatic loss because of cannibalism by endodermal cells during embryonic development, and dramatic gain because of material from the distal germline and the intestine during oocyte maturation.

In summary, further development of information theory that incorporates the above considerations for application to biological systems coupled with experimental manipulation in well-characterized systems like *C. elegans* could enable rigorous analyses of cell codes (see Ref. [110] for additional perspectives on information in biology).




*Cell code*: the information for making an organism that is stored in single cells and transmitted from one generation to the next by interdependent replicating and cycling molecules. A) A cell accrues components guided by the action of preexisting components on DNA sequence ($C_0$, $C_1$, $C_2$, ...). New components made at each moment ($t$) can interact with, modify, and/or be modified by (•) preexisting components, which could have changed since they were made (e.g., $C_{0\to2}$ and $C_{1\to2}$ at $t=2$). This process extends to external inputs the cell interacts with or acquires independent of its DNA ($E_0$, $E_1$, $E_2$, ...). At a given time ($t=n$), the cellular content is equal to the accrued cellular contents ($C_{past\to n}$) that have changed through interactions with the accrued external input ($E_{past\to n}$). B) DNA sequence ($DNA_{seq}$) and a specific arrangement of components accrued over time ($C_{past\to n}$•$E_{past\to n}$) that are recreated in single cells that begin successive generations of an organism (Gen $x$, Gen $x+1$) define the cell code of that organism. C) Transmission of the cell code may be fruitfully analyzable as information transfer through *interdependent* channels for the transmission of cellular components (DNA, proteins, RNAs, spatial arrangements, etc.). Changes that occur during transmission could be thought of as being akin to noise in a communication channel and could impact the transmission of DNA sequence information (e.g., mutations and meiosis) or other channels (e.g., interaction with somatic cells and the environment). D) The well-characterized *C. elegans* germline, along which its cell code is transmitted from one generation to the next, invites examination of systematic and stochastic changes that occur during such transmission (see text for details). Barring the DNA sequence, all other aspects of the cell code (including replicating and cycling stores of information) are transmitted via unknown mechanisms.

by signals, uptake of nucleic acids, etc.) accrued from additional sources such as the microbiome or maternal circulation during development after the single-cell bottleneck that could also vary across generations.

Modifications to the cell code can alter its genomic sequence ($DNA_{seq}$, see Box 3) or three-dimensional arrangement of all components ($C_{past\to n}$•$E_{past\to n}$, see Box 3) within the cell. Changes to the DNA sequence can persist for many generations and are

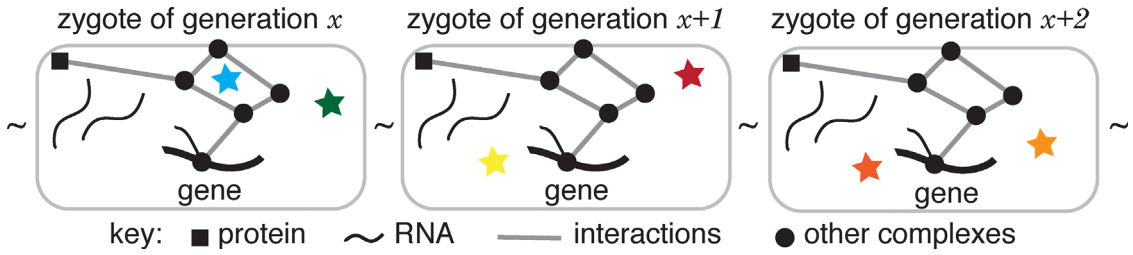



**Figure 2.** Nature of the cell code. Part of the cell code from the perspective of a single gene (thick black line), its RNA (curved lines), and its protein (square) is depicted. Complexes of other proteins/RNAs/lipids, etc. (circles) potentially localized in specific places within the cell and interactions between them (gray lines) that can impact the gene and that are reproduced in every generation are part of the cell code. Some components (stars) in the zygote may not be inherited in successive generations and compensatory changes in the rest of the zygote may permit different configurations for this gene while preserving the overall cell code of the organism.

indeed the best-studied changes to the cell code. These can include both random mutations and changes derived from experience (e.g., CRISPR-Cas system of antiviral immunity in bacteria[40]). But, changes to the cell code that do not alter DNA sequence also have the potential to persist for many generations. In fact, one of the earliest mutants ever described – in 1744 – was found to not alter the DNA sequence. This variant of the toadflax *Linaria vulgaris* called *peloria* (greek πελωρ, monster)[41,42] arises at ≈1% frequency in each generation[43] and is correlated with methylation but not sequence changes at the *Lcyc* locus.[44] Such inheritance of changes that do not alter DNA sequence and yet affect how the DNA sequence is used in any generation can be viewed as the influence of organismal history. The controlling influence of past cellular components on the production of future components from DNA is powerfully illustrated in ciliates that develop using RNA from one generation to splice or unscramble germline DNA sequences in the next generation. In these ciliates, such as *Tetrahymena* and *Oxytricha*, changes in RNA introduced in one generation can alter how the DNA is spliced or unscrambled in many subsequent generations.[45,46] In other organisms, extensively studied phenomena like paramutation (see Ref. [47] and reviewed in Ref. [48]) and RNA interference[49] that cause persistent silencing of genes also illustrate the persistence of non-genetic information. While the precise mechanisms for how the gene silencing information is transmitted across generations are still being worked out, persistent chromatin modifications and amplified RNAs have emerged as possibilities (reviewed in Ref. [50]). Transmission for many generations is likely to require inheritance of molecules containing the information for silencing a gene for one or a few generations combined with periodic replication of molecules to reinforce this information. For example, an ancestral event such as exposure to double-stranded RNA during RNA interference could trigger production of chromatin modifications in each generation based on instructions held in RNA. All that is needed to make this information stable for many generations is transmission of chromatin and/or RNA, even for just one generation, followed by reproduction in every generation. Crucially, the resulting permanent changes to the cell code do not involve changes in the DNA sequence. Other mechanisms for changing the cell code without mutating DNA include the transmission of prions,[51] where alternative folding states of proteins that can template recreation of similar states are transmitted across generations.

When complex organisms, like humans, that can make tools and artifacts by interacting with both the environment and other organisms are considered, there is no end to the longevity of ancestral information and its influence on physiology. For example, the spread of ideas within a culture (i.e., a meme[52]) can have a profound influence on human behavior and can drive physiological changes as evidenced in dietary practices. Thus, the evolution of such organisms is shaped by much more than the molecules present in their cells.[53] In fact, contributions by many mechanisms that transmit information across generations can be incorporated into a consideration of how organisms evolve[54] and discerning the relative importance of these mechanisms requires carefully controlled experiments.

Even if only the transmission of biological material is considered, the precise way in which information is transmitted across generations is unclear. While DNA sequence ($DNA_{seq}$) in a zygote is from the immediate parents, the provenance of the information that specifies the arrangement of all molecules ($C_{past\rightarrow n} \bullet E_{past\rightarrow n}$) in a zygote is less clear. If life originated once on earth as suggested by the commonalities among extant organisms, then every zygote can be traced back to the first cell by descent. This initial cell must have accrued complexities that were transmitted from one generation to the next. Therefore, although the molecules present in the fertilized zygote are inherited from the gametes of the immediate parents, it is conceivable that their amounts and/or arrangement changed in an ancestral generation, and this change has since been propagated without modification. Nevertheless, there have only been a few cases of clear evidence for such non-genetic changes that persist for many generations (see Ref. [55] for review of early work). This paucity could reflect limited experimentation, incompatibility of the changes with the perpetuation of life, or mechanisms that oppose such persistence.

## 5. Forces That Oppose Variation Preserve the Cell Code

The development of an organism in a given environment follows a path that reflects the programmed constraints imposed by epigenetic control systems within cells and cell-cell interactions in that organism. This constrained path was referred to as "chreod" ("necessary path" from greek roots χρη, it is necessary, and όδ∘ζ, path[56]) and has been given modern form within the framework of dynamical systems theory.[57] Such developmental constraints have been recognized as one of the forces that oppose change in organisms during evolution.[58]

Some studies show correlation of certain molecular changes in an animal with environmental or dietary changes in parents or recent ancestors (e.g., Ref. [59–63]) and molecular changes such as histone methylation that occur within germ cells can cause effects that persist for a few generations (e.g., Ref. [64,65]). These instances of heritable changes to non-genetic aspects of the cell code urge consideration of the possible ancestral origins of disease and the impact of medical intervention on descendants. However, loss or erasure of such chemical changes after a few generations reflects a homeostatic return to a prior cell code (**Figure 3**). For example, during early mammalian development parental DNA methylation is erased at most sites and new modifications are added.[66] Because this happens in every generation, the information for adding these new modifications must be passed from one generation to the next as an aspect (molecule and/or arrangement) of the cell code. This developmental reprogramming thus preserves the cell code and opposes transgenerational epigenetic changes. Nevertheless, cases of persistent non-genetic changes — some lasting for tens to hundreds of generations (e.g., Ref. [67–78]) — provide us with opportunities to analyze how non-genetic aspects of the cell code can be reconfigured.

Together, these considerations inform attempts to identify past events that have shaped the evolution of organisms and future attempts to synthesize organisms.

## 6. To Make Life We Need to Know the Cell Code

Considering what we would need to know to make an organism from raw materials reveals the extent of our ignorance of the cell code. In recent years, we have begun to remove, replace, augment, or modify aspects of the cell code. In particular, techniques for manipulating DNA sequence have advanced to the point that practically any change in DNA sequence can be made using genome engineering (reviewed in Ref. [79]) and even the entire genome of a simple organism can be replaced with a synthetic genome.[80] We can effectively transfer genetic material to avoid mutant mitochondria,[81] augment the genetic code,[82] and even use gene drives to change wild populations.[83] All these efforts are akin to modifying a pre-existing machine by tinkering with or replacing parts without fully understanding how the machine is put together. As with complex machinery, such manipulations without deep understanding can be dangerous. Additionally, all manipulations of life warrant careful ethical considerations (e.g., Ref. [84]). To make a machine from raw materials, however, we need to know the *entire* design. Because organisms build themselves, assembling the cell code of an organism could be sufficient to make an organism. Although an assembly of chemicals that display characteristics of life has been imagined in theory (e.g., The Chemoton Theory[4]), the enormity of this challenge in practice is clear in our struggle to use raw materials to make a rudimentary cell that perpetuates itself.[85,86] Nevertheless, we have begun to coax pre-existing cells into making complex parts of organisms in vitro — for example, eye[87] and brain.[88] Eventually making an entire complex organism that reproduces will require first discovering the cell code of that organism and then determining the simplest versions of that code. The "endless forms most beautiful and most wonderful"[39] found in nature each have a different cell code. Such organisms sculpted by evolution, however, are unlikely to already have the minimal essential cell code because evolution tinkers with what exists[89] but anticipates poorly. Our current understanding suggests that evolution proceeds through measure and counter-measure, including nonadaptive processes

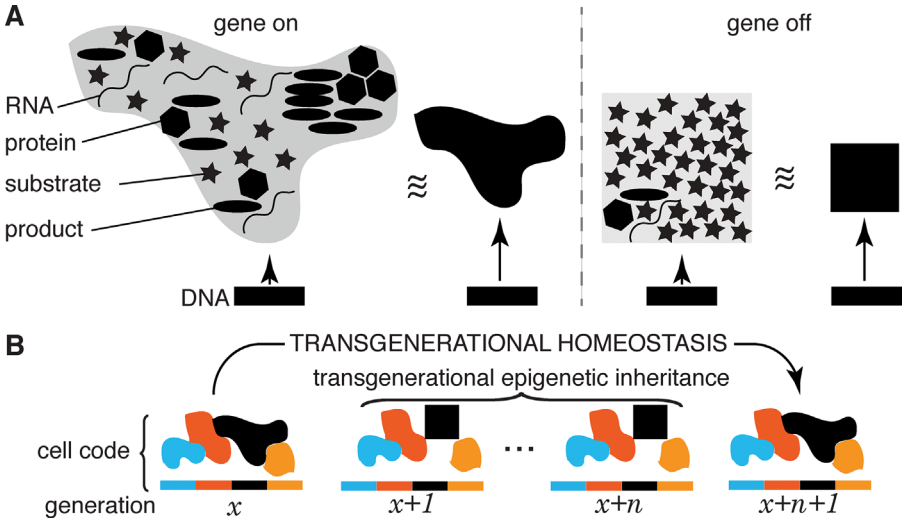



**Figure 3.** The balance between transgenerational homeostasis and transgenerational epigenetic inheritance determines heritable non-genetic variation. A) Schematic illustrating impact on cellular components of high (gene on) versus low (gene off) expression of a gene encoding an enzyme. B) Non-genetic changes to a cell code (e.g., down regulation of gene expression) that occur during one generation ($x$, say) can persist through transgenerational epigenetic inheritance for many generations (until $x+n$, say) and then return to ancestral states because of transgenerational homeostasis. Cases of indefinite persistence of change (i.e., large $n$) could reflect establishment of a new set point for transgenerational homeostasis. Non-genetic changes that persist for a few or for hundreds of generations have been detected in some organisms,[67–78] but the reasons for susceptibility to change, persistence of change, or homeostatic return to ancestral states after a period of change are unknown.

(e.g., Ref. [90,91]), resulting in a cellular architecture with superfluous aspects.[92] Thus, organisms found in nature have layers of post hoc regulation that can obscure essential design principles. Understanding such principles can enable the design of organisms free of the historically contingent tinkering of evolution.

## 7. Approaches to Decipher the Cell Code

Discovering the cell code of an organism minimally requires comparison of features in zygotes of successive generations (Figure 2). Our increasing ability to perform molecular studies on single cells[93] makes this approach reasonable. Parental effect mutants can be used to identify potential components of the cell code that are maternally or paternally deposited into the zygote. Such mutants may also identify signaling from one generation to the next. However, the arrangement of these molecules in the zygote is more difficult to discover and would require systematic cell biological and biochemical analyses. After these components are identified and their arrangements are discovered, the zygotes of successive generations can be compared.

### 7.1. Perturbation Studies

Penetrating insights often require perturbation of the system and not mere observation. Past perturbation studies of development and reproduction have largely focused on the analysis of essential genes that impact viability or fertility, respectively. Consequently, the effects of a perturbation are typically only analyzed during a limited period of the life cycle. Furthermore, defects in an essential gene (required for development, say) would kill the organism precluding examination of the same stage in successive generations. Thus, when we interfere with an essential gene or process in one generation, we lose the ability to see its impact on the next generation because the intervening organism is affected. In other words, we cannot know if an essential gene or process is required to make the organism or to reproduce the cell code or both. Studies using viable and fertile mutants on the other hand permit examination of the zygote in successive generations. But, careful subsequent analysis will be needed to separate defects in the mere making of the organism from perturbations of the cell code. For example, a mutation in DNA that changes a residue in hemoglobin alters genome sequence in the cell code and affects the structure of the protein made in red blood cells, but likely does not affect non-genetic aspects of the cell code or its propagation via the germline. However, a mutation in DNA that changes a protein that is reproducibly present in the zygote or the germline changes not only the genome sequence in the cell code but potentially also non-genetic aspects of the cell code and/or its propagation. Finally, a non-genetic change (e.g., chromatin modification or addition of double-stranded RNA) that causes gene silencing that persists for many generations alters non-genetic aspects of the cell code and potentially its propagation without changing the genome sequence.

Nuclear transplantation experiments were essential for the realization that most cells in an organism retain the same genome.[94,95] A similar approach could be used to discover all the cells that retain the ability to use the DNA genome to generate the entire organism, that is, to discover all the cells that have the non-genetic aspects (molecules and arrangement) of the cell code. Every cell that can be induced to generate the entire organism is expected to either have, or be able to reconfigure its contents to create, the entire cell code. For sexually reproducing organisms, minimally every cell that is continuous within the female germline from one zygote to the next is expected to have all the non-genetic aspects of the cell code except those obtained from the male gamete upon fertilization. Conversely, every cell that is continuous within the male germline from one zygote to the next is expected to have all the non-genetic aspects of the cell code except those obtained from the female gamete upon fertilization. It will be interesting to discover the somatic cells that retain non-genetic aspects of the cell code despite differentiation during normal development and to discover how the cell code is reduced from being present in its entirety in the zygote to the portion in each gamete awaiting union upon fertilization.

The ability to make random changes in DNA and examine its consequences – forward genetics – and more recently to turn off any gene through RNA interference – reverse genetics – were crucial for correlating changes in the genome with changes in the organism. To correlate changes in non-genetic components of the cell code with changes in the organism, we additionally need forward epigenetics and reverse epigenetics. While we do not yet have a way to perform forward epigenetics, reverse epigenetics has begun.[96] For example, a guide RNA and the Cas9 enzyme fused to a histone modifier can be used to target chromatin modifications to histones located in a specific region of the genome (e.g., Ref. [97]). Such a manipulation could be performed in one generation and its impact analyzed in subsequent generations to determine whether the manipulation altered non-genetic aspects of the cell code.

Although the total information that is transmitted from one generation to the next is high-dimensional, to make rapid progress in understanding the principles of the cell code, we are likely to also benefit from focusing on one or a few genes. Analyzing a few genes, their protein and/or RNA products, and the factors that impact their recreation in successive generations will allow us to ask specific questions and generate a preliminary understanding of the cell code. This focused approach could be an effective complement to comprehensive approaches that compare many components of zygotes in successive generations in response to experimental perturbation. In support of this traditional approach, the basic principles of gene regulation were worked out through sustained effort on a few genes – for example, the *lac* operon.[98]

### 7.2. Tracer Studies

An alternative to perturbation studies that overcomes the impasse of lethality or sterility when the cell code is disturbed is the use of tracer studies. To illustrate this approach, imagine you wanted to discover the extent of the circulatory system of an organism. A perturbation study could involve making cuts throughout its body and examining if blood spurts out. A tracer

study could involve injecting an inert molecule that permeates the entire circulatory system and then imaging that molecule. Therefore, taking the tracer approach to examine the cell code of an organism, we could insert benign sequences (e.g., encoding a fluorescent protein) into its genome, add or remove various regulatory features, and examine if these changes have consequences that persist across generations. In support of the power of this approach, the deciphering of the genetic code began with a "tracer" study where one RNA sequence (poly-U) was used to synthesize one peptide sequence (poly-Phe) in an in vitro translation system.[99]

### 7.3. Different Organisms Warrant Different Considerations

In organisms that have well-differentiated somatic and germ cells,[100] it will be interesting to discover the extent of somatic influence on the perpetuation of the cell code via the germline. A germ cell from one animal could be transplanted into another[101] to evaluate the extent of somatic influence on the cell code. Similarly, the circulatory system of two animals could be joined together (reviewed in Ref. [102]) to examine the effect of secreted material from the cells of one animal on the germline of another. Evidence for such influence of secreted material is provided by intriguing observations like the entry of double-stranded RNA from neurons into the germline in the worm *C. elegans*, resulting in gene silencing that persists for more than 25 generations.[73] Analyzing somatic influences in different environments could reveal mechanisms, if any, by which the environment could alter the cell code in different organisms.

Simpler organisms likely have simpler cell codes. Single-celled organisms like the marine green alga *Ostreococcus tauri*[103] or bacteria could be chosen as model systems to discover the simplest cell codes. Simpler still are organisms that result from efforts to generate cells that have a minimal genome.[37,38] While the analysis of these organisms would reveal the logic of cell codes for single-celled organisms, it is conceivable that the need for differentiation in complex animals and plants results in fundamentally different strategies for the assembly and reproduction of their cell codes.

## 8. Conclusion

The sequence of a genome can be used to identify an organism. However, the genome sequence is not sufficient information for making that organism. To make an organism we need to know its cell code, which is the evolving arrangement of molecules, including the genome, that is nearly reproduced in every generation. Building on more than a century of work in biology, we can now begin to decipher the cell code of an organism by analyzing single cells that start successive generations.

## Acknowledgements

The ideas presented here were honed through long conversations with my friend Katerina Ragkousi, through lectures to students of a course on Ignorance run by José Feijo, and through a faculty chalk talk to colleagues at the University of Maryland. I thank Steve Wolniak for initially pointing me to relevant books on cell biology; Karen Carleton, David Mosser, Zhongchi Liu, Caren Chang, Norma Andrews, José Feijo, Wenxia Song, Carlos Machado, Tom Kocher, Charles Delwiche, William Snell, Steve Wolniak, Sougata Roy, Geraldine Seydoux, Kevin O'Connell, Andy Golden, John Hanover, Victor Ambros, Marie-Anne Felix, Pierre-Emmanuel Jabin, and anonymous reviewers for comments on the manuscript; all past and present members of the Jose lab for lively discussions; and my friend John Horner for early support when these ideas were in their infancy. I apologize to any colleagues whose work could not be highlighted in this short article. Research in the author's lab is supported through funds from the University of Maryland and from the NIH (R01GM111457).

## Conflict of Interest

The author has declared no conflict of interest.

## Keywords

development, epigenetics, evolution, homeostasis, synthetic biology, systems biology, transgenerational inheritance

Received: September 1, 2017
Revised: January 6, 2018
Published online:

[1] B. Commoner, *Nature* **1964**, *202*, 960.
[2] F. J. Dyson, *J. Mol. Evol*. **1982**, *18*, 344.
[3] J. Maynard Smith, *J. Theor. Biol*. **1990**, *143*, 41.
[4] T. Ganti, *The principles of life*. Oxford University Press, Oxford (originally published in 1971) **2003**, p. 4.
[5] F. M. Harold, *Microbiol. Mol. Biol. Rev*. **2005**, *69*, 544.
[6] M. Ptashne, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7101.
[7] M. Thanbichler, L. Shapiro, *Nat. Rev. Microbiol*. **2008**, *6*, 28.
[8] S. W. Englander, L. Mayne, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15873.
[9] L. E. O'Brien, T. S. Jou, A. L. Pollack, Q. Zhang, S. H. Hansen, P. Yurchenco, K. E. Mostov, *Nat. Cell Biol*. **2001**, *3*, 831.
[10] C. Mereschkowsky, *Biol. Centralbl*. **1905**, *25*, 593.
[11] L. Sagan, *J. Theoret. Biol*. **1967**, *14*, 255.
[12] S. M. Rafelski, W. F. Marshall, *Nat. Rev. Mol. Cell Biol*. **2008**, *9*, 593.
[13] D. L. Nanney, *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 712.
[14] J. Monod, F. Jacob, *Cold Spring Harbor Symp. Quant. Biol*. **1961**, *26*, 389.
[15] E. Mayr, *Science* **1961**, *134*, 1501.
[16] S. Kauffman, *Nature* **1969**, *224*, 177.
[17] M. Levine, E. H. Davidson, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 4936.
[18] K. Sander, *Roux's Arch. Dev. Biol*. **1991**, *200*, 1.
[19] E. B. Wilson, *The Cell in Development and Inheritance*. The Macmillian Company, New York **1896**.
[20] T. Boveri, *Neu Folge* **1902**, *35*, 67. Translated by Gluecksohn-Waelsch S, (B. H. Willer and J. Oppenheimer, eds). Prentiss-Hall, Inc., Englewood Cliffs, NJ, 1964.
[21] T. Boveri, *Die Organismen als Historche Wesen (Organisms as historical beings)*. Kgl. Universitats Druckerei von H. Sturtz, Wurzburg **1906**.
[22] E. G. Conklin, *Science* **1908**, *27*, 89.
[23] C. H. Waddington, *New Patterns in Genetics and Development*. Columbia University Press, New York **1962**.
[24] C. H. Waddington, *Endeavour* **1942**, *1*, 18.
[25] C. H. Waddington, *Towards a Theoretical Biology, Vol. 1, Prolegomena*. International Union of Biological Sciences & Edinburgh University Press, Edinburgh **1968**, pp. 1.

[26] J. Huxley, *Evolution: The Modern Synthesis — the definitive edition*. MIT Press, Cambridge (originally published in1942 London: Allen and Unwin) **2010**, p. 555.
[27] E. J. A. Hubbard, D. Greenstein, *WormBook*. The *C. elegans* Research Community, WormBook, **2015**, https://doi.org/10.1895/wormbook.1.18.1, http://www.wormbook.org.
[28] Y. Abdu, C. Maniscalco, J. M. Heddleston, T. L. Chew, J. Nance, *Nat. Cell Biol*. **2016**, *18*, 1302.
[29] U. Wolke, E. A. Jezuit, J. R. Priess, *Development* **2007**, *134*, 2227.
[30] B. Grant, D. Hirsh, *Mol. Biol. Cell*. **1999**, *10*, 4311.
[31] J. Marré, E. C. Traver, A. M. Jose, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12496.
[32] J. Marré, A. Jose, *Mol. Reprod. Dev*. **2017**, *84*, 283.
[33] E. Wang, C. P. Hunter, *Genetics* **2017**, *207*, 547.
[34] E. D. Salmon, *J. Cell Biol*. **1975**, *65*, 603.
[35] S. B. Weiss, L. Gladstone, *J. Am. Chem. Soc*. **1959**, *81*, 4118.
[36] F. R. Lillie, *J. Morphol*. **1896**, *12*, 239.
[37] C. A. Hutchison, 3rd, R. Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z. Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y. Suzuki, B. Tsvetanova, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, J. C. Venter, *Science* **2016**, *351*, aad6253.
[38] J. I. Glass, C. Merryman, K. S. Wise, C. A. Hutchison, 3rd, H. O. Smith, *Cold Spring Harb. Perspect. Biol*. **2017**, *9*, a023861.
[39] C. Darwin, *On the Origin of Species*, 1st edition. John Murray, London **1859**.
[40] R. Barrangou, *Curr. Opin. Immunol*. **2015**, *32*, 36.
[41] C. Linnaeus, D. Rudberg, *De Peloria*. Diss. Ac. Amoenitates Academicae III, Uppsala **1744**.
[42] A. Gustafsson, *Theor. Appl. Genet*. **1979**, *54*, 241.
[43] H. DeVries, *Species and Varieties: Their origin and mutation*, 2nd edition. The Open Court Publishing Company, London **1906**, p. 475.
[44] P. Cubas, C. Vincent, E. Coen, *Nature* **1999**, *401*, 157.
[45] J. R. Bracht, W. Fang, A. D. Goldman, E. Dolzhenko, E. M. Stein, L. F. Landweber, *Cell* **2013**, *152*, 406.
[46] D. L. Chalker, E. Meyer, K. Mochizuki, *Cold Spring Harb. Perspect. Biol*. **2013**, *5*, a017764.
[47] R. A. Brink, *Genetics* **1956**, *41*, 872.
[48] J. B. Hollick, *Nat. Rev. Genet*. **2017**, *18*, 5.
[49] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, *Nature* **1998**, *391*, 806.
[50] C. H. Rankin, *J. Exp. Biol*. **2015**, *218*, 41.
[51] R. Halfmann, S. Lindquist, *Science* **2010**, *330*, 629.
[52] R. Dawkins, *The Selfish Gene*. Oxford University Press, Oxford **1989**, p. 192.
[53] E. Jablonka, M. L. Lamb, *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. The MIT Press, Cambridge **2005**.
[54] O. Rivoire, S. Leibler, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1940.
[55] E. Jablonka, G. Raz, *Quant. Rev. Biol*. **2009**, *84*, 131.
[56] C. H. Waddington, *The strategy of the genes*. George Allen & Unwin Ltd, New York **1957**, p. 32.
[57] A. O. Pisco, A. F. d'Herouel, S. Huang, *BioRxiv* **2016**, 1. Online on May 12.
[58] S. J. Gould, N. Eldredge, *Nature* **1993**, *366*, 223.
[59] A. B. Rodgers, C. P. Morgan, S. L. Bronson, S. Revello, T. L. Bale, *J. Neurosci*. **2013**, *33*, 9003.
[60] O. Rechavi, L. Houri-Ze'evi, S. Anava, W. S. Goh, S. Y. Kerk, G. J. Hannon, O. Hobert, *Cell* **2014**, *158*, 277.
[61] U. Sharma, C. C. Conine, J. M. Shea, A. Boskovic, A. G. Derr, X. Y. Bing, C. Belleannee, A. Kucukural, R. W. Serra, F. Sun, L. Song, B. R. Carone, E. P. Ricci, X. Z. Li, L. Fauquier, M. J. Moore, R. Sullivan, C. C. Mello, M. Garber, O. J. Rando, *Science* **2016**, *351*, 391.
[62] Q. Chen, M. Yan, Z. Cao, X. Li, Y. Zhang, J. Shi, G. H. Feng, H. Peng, X. Zhang, Y. Zhang, J. Qian, E. Duan, Q. Zhai, Q. Zhou, *Science* **2016**, *351*, 397.
[63] A. Klosin, E. Casas, C. Hidalgo-Carcedo, T. Vavouri, B. Lehner, *Science* **2017**, *356*, 320.
[64] E. L. Greer, T. J. Maures, D. Ucar, A. G. Hauswirth, E. Mancini, J. P. Lim, B. A. Benayoun, Y. Shi, A. Brunet, *Nature* **2011**, *479*, 365.
[65] K. Siklenka, S. Erkek, M. Godmann, R. Lambrot, S. McGraw, C. Lafleur, T. Cohen, J. Xia, M. Suderman, M. Hallett, J. Trasler, A. H. Peters, S. Kimmins, *Science* **2015**, *350*, aab2006.
[66] S. Feng, S. E. Jacobsen, W. Reik, *Science* **2010**, *330*, 622.
[67] N. L. Vastenhouw, K. Brunschwig, K. L. Okihara, F. Müller, M. Tijsterman, R. H. Plasterk, *Nature* **2006**, *442*, 882.
[68] A. Ashe, A. Sapetschnig, E. M. Weick, J. Mitchell, M. P. Bagijn, A. C. Cording, A. L. Doebley, L. D. Goldstein, N. J. Lehrbach, J. Le Pen, G. Pintacuda, A. Sakaguchi, P. Sarkies, S. Ahmed, E. A. Miska, *Cell* **2012**, *150*, 88.
[69] B. A. Buckley, K. B. Burkhart, S. G. Gu, G. Spracklin, A. Kershner, H. Fritz, J. Kimble, A. Fire, S. Kennedy, *Nature* **2012**, *489*, 447.
[70] M. J. Luteijn, P. van Bergeijk, L. J. Kaaij, M. V. Almeida, E. F. Roovers, E. Berezikov, R. F. Ketting, *EMBO J*. **2012**, *31*, 3422.
[71] M. Shirayama, M. Seth, H. C. Lee, W. Gu, T. Ishidate, D. Conte, Jr, C. C. Mello, *Cell* **2012**, *150*, 65.
[72] A. de Vanssay, A. L. Bougé, A. Boivin, C. Hermant, L. Teysset, V. Delmarre, C. Antoniewski, S. Ronsseray, *Nature* **2012**, *490*, 112.
[73] S. Devanapally, S. Ravikumar, A. M. Jose, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2133.
[74] L. E. Leopold, B. N. Heestand, S. Seong, L. Shtessel, S. Ahmed, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2667.
[75] O. Minkina, C. P. Hunter, *Mol. Cell* **2017**, *65*, 659.
[76] I. Lev, U. Seroussi, H. Gingold, R. Bril, S. Anava, O. Rechavi, *Curr. Biol*. **2017**, *27*, 1138.
[77] F. Ciabrelli, F. Comoglio, S. Fellous, B. Bonev, M. Ninova, Q. Szabo, A. Xuéreb, C. Klopp, A. Aravin, R. Paro, F. Bantignies, G. Cavalli, *Nat. Genet*. **2017**, *49*, 876.
[78] S. Devanapally, S. Allgood, A. M. Jose, *BioRxiv* **2017**, 1. Online on Jun 11.
[79] A. C. Komor, A. H. Badran, D. R. Liu, *Cell* **2017**, *168*, 20.
[80] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z. Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, 3rd, H. O. Smith, J. C. Venter, *Science* **2010**, *329*, 52.
[81] M. Tachibana, M. Sparman, H. Sritanaudomchai, H. Ma, L. Clepper, J. Woodward, Y. Li, C. Ramsey, O. Kolotushkina, S. Mitalipov, *Nature* **2009**, *461*, 367. Erratum in: Nature 2014, 516, 276.
[82] N. Ostrov, M. Landon, M. Guell, G. Kuznetsov, J. Teramoto, N. Cervantes, M. Zhou, K. Singh, M. G. Napolitano, M. Moosburner, E. Shrock, B. W. Pruitt, N. Conway, D. B. Goodman, C. L. Gardner, G. Tyree, A. Gonzales, B. L. Wanner, J. E. Norville, M. J. Lajoie, G. M. Church, *Science* **2016**, *353*, 819.
[83] V. M. Gantz, E. Bier, *Science* **2015**, *348*, 442.
[84] D. Baltimore, P. Berg, M. Botchan, D. Carroll, R. A. Charo, G. Church, J. E. Corn, G. Q. Daley, J. A. Doudna, M. Fenner, H. T. Greely, M. Jinek, G. S. Martin, E. Penhoet, J. Puck, S. H. Sternberg, J. S. Weissman, K. R. Yamamoto, *Science* **2015**, *348*, 36.
[85] J. C. Blain, J. W. Szostak, *Annu. Rev. Biochem*. **2014**, *83*, 615.
[86] P. Schwille, *J. Cell Biol*. **2015**, *210*, 687.
[87] M. Eiraku, N. Takata, H. Ishibashi, M. Kawada, E. Sakakura, S. Okuda, K. Sekiguchi, T. Adachi, Y. Sasai, *Nature* **2011**, *472*, 51.
[88] M. A. Lancaster, M. Renner, C. A. Martin, D. Wenzel, L. S. Bicknell, M. E. Hurles, T. Homfray, J. M. Penninger, A. P. Jackson, J. A. Knoblich, *Nature* **2013**, *501*, 373.
[89] F. Jacob, *Science* **1977**, *196*, 1161.

[90] M. Lynch, *Proc. Natl. Acad. Sci. USA* **2007**, *1*, 8597.
[91] H. D. Madhani, *Cell* **2013**, *155*, 744.
[92] S. J. Gould, R. C. Lewontin, *Proc. R. Soc. Lond. B Biol. Sci.* **1979**, *205*, 581.
[93] K. Tsioris, A. J. Torres, T. B. Douce, J. C. Love, *Annu. Rev. Chem. Biomol. Eng.* **2014**, *5*, 455.
[94] R. Briggs, T. J. King, *Proc. Natl. Acad. Sci. USA* **1952**, *38*, 455.
[95] J. B. Gurdon, T. R. Elsdale, M. Fischberg, *Nature* **1958**, *182*, 64.
[96] M. Park, A. J. Keung, A. S. Khalil, *Genome Biol.* **2016**, *17*, 183.
[97] I. B. Hilton, A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy, C. A. Gersbach, *Nat. Biotechnol.* **2015**, *33*, 510.
[98] F. Jacob, J. Monod, *J. Mol. Biol.* **1961**, *3*, 318.
[99] M. W. Nirenberg, J. H. Matthaei, *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1588.
[100] C. G. Extavour, M. Akam, *Development* **2003**, *130*, 5869.
[101] R. L. Brinster, J. W. Zimmermann, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11298.
[102] A. Eggel, T. Wyss-Coray, *Swiss Med. Wkly.* **2014**, *144*, w13914.
[103] E. Derelle, C. Ferraz, S. Rombauts, P. Rouzé, A. Z. Worden, S. Robbens, F. Partensky, S. Degroeve, S. Echeynié, R. Cooke, Y. Saeys, J. Wuyts, K. Jabbari, C. Bowler, O. Panaud, B. Piégu, S. G. Ball, J. P. Ral, F. Y. Bouget, G. Piganeau, B. De Baets, A. Picard, M. Delseny, J. Demaille, Y. Van de Peer, H. Moreau, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11647.
[104] J. Beisson, T. M. Sonneborn, *Proc. Natl. Acad. Sci. USA* **1965**, *53*, 275.
[105] J. S. Torday, *Biology (Basel)*. **2015**, *4*, 573.
[106] E. F. Phillips, *Proc. Amer. Phil. Soc.* **1903**, *42*, 275.
[107] O. W. Olsen, *Animal Parasites: Their Life Cycles and Ecology*. Dover Publications, Mineola, New York **1986**.
[108] K. D. Birnbaum, A. Sanchez-Alvarado, *Cell* **2008**, *132*, 697.
[109] C. Shannon, *Bell Syst. Tech. J.* **1948**, *27*, 379.
[110] J. H. E. Cartwright, S. Giannerini, D. L. Gonzalez, Philos. Trans. Royal Soc. A 2016, 374. a) M. Barbieri, *Philos. Trans. Royal Soc. A* **2016**, *374*, 20150060; b) E. V. Koonin, *Philos. Trans. Royal Soc. A* **2016**, *374*, 20150065; c) P. R. Wills, *Philos. Trans. Royal Soc. A* **2016**, *374*, 20150066; d) P. Ball, *Philos. Trans. Royal Soc. A* **2016**, *374*, 20150072; e) M. Barbieri, *Philos. Trans. Royal Soc. A* **2016**, *374*, 20150148; f) C. Adami, *Philos. Trans. Royal Soc. A* **2016**, *374*, 20150230.