Biol Theory (2013) 8:221–231
DOI 10.1007/s13752-013-0131-1

THEMATIC ISSUE ARTICLE: EMERGENCE OF SHAPE

The Origin of Metazoa: An Algorithmic View of Life

Rafaele Di Giacomo · Jeffrey H. Schwartz · Bruno Maresca

Received: 24 February 2013 / Accepted: 25 February 2013 / Published online: 10 July 2013

Abstract We propose that the sudden emergence of metazoans during the Cambrian was due to the appearance of a complex genome architecture that was capable of computing. In turn, this made defining recursive functions possible. The underlying molecular changes that occurred in tandem were driven by the increased probability of maintaining duplicated DNA fragments in the metazoan genome. In our model, an increase in telomeric units, in conjunction with a telomerase-negative state and consequent telomere shortening, generated a reference point equivalent to a non-reversible counting mechanism.

Keywords Development · Turing machine · Göedel's theorems

> Death is very likely the single best invention of life.
> —Steve Jobs, Stanford University, 2005

Porifera (sponges) are generally believed to represent a basal extant metazoan phylum and its members the most primitive multicellular organisms (Kruse et al. 2000). However, while poriferans are multicellular, unlike cnidarians, they are not true metazoans because they express neither a defined differentiated state nor organismal symmetry. Instead, the "organism" is merely a mass of cells that can reproduce asexually and sexually even though it lacks distinct gonads. Further, while poriferans produce different cell types, such as those lining the "mass's" external and internal surfaces, these do not differentiate into true tissues, and most are totipotent (i.e., capable of changing form and function). Thus while a sponge is a multicellular entity, it remains at the unicellular level of complexity (Brusca and Brusca 2003). Further, a sponge's shape is determined primarily by environmental factors (e.g., water flow dynamics).

In the demosponge *Suberites domuncula*, telomerase-positive cells constitutively express telomerase and thus have the potential for unlimited cell division. Cell lineages are, therefore, potentially immortal. Cells are triggered to apoptosis and death when single cells lose cell–cell and/or cell–matrix contact, separate from the cell mass, and become telomerase negative, which leads to telomere shortening (Koziol et al. 1998). Also in *S. domuncula* SDLAGL, which is a longevity assurance-like gene coding for a 330 aa long polypeptide, is involved in the shift from telomerase-positive (proliferating) cells to telomerase-negative (non-proliferating, telomere-shortening, apoptotic) cells. The simultaneous manifestation of controlled cell death and deactivation of telomerase with concomitant telomere erosion in non-proliferating single sponge cells suggests that multiplication of DNA fragments (such as telomeric repetitive units) could increase genome size because amplified telomeric repeats would extend cell-lineage longevity (i.e., postpone cell-lineage death) by increasing the number of viable cell generations prior to reaching the lowest limit of telomere length that would enable chromosome replication.

R. Di Giacomo
Department of Electronic and Informatics Engineering, University of Salerno, Salerno, Italy

J. H. Schwartz
Department of Anthropology and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA

B. Maresca (✉)
Division of BioMedicine, Department of Pharmacy, University of Salerno, Salerno, Italy
e-mail: bmaresca@unisa.it

As observed in *S. domuncula*, when an isolated telomerase-negative cell divides, one daughter cell retains full-length chromosomes and telomeres, while the other receives shorter telomeres and, thus, shorter chromosomes, because the DNA lagging strand is discontinuously synthesized. Further in *S. domuncula*, when daughter cells divide, granddaughter cells with shortened telomeres enter apoptosis and die (Fig. 1). A similar mechanism of asymmetrical cell/chromosome division could have led to the emergence of a simple, two-celled "organism," in which one cell lineage dies quickly while another perpetuates a "germ line." It is, therefore, not unreasonable to hypothesize that the first, two-celled but cellularly "differentiated" organism would have been genetically telomerase negative and thus very different from its parent (presumably something like a colonial sponge cell) and would have transmitted its novelties of linear-chromosomal DNA with repeat segments and telomere regulation to its descendants, which would remain isolated cells.

Based on the poriferan evidence, we hypothesize that duplication of telomeric units and incremental shortening of telomeres within a few cell divisions made possible an increase in both genome size and gene number. Since telomeric regions fold over and mask subtelomeric DNA (Dubrana et al. 2001), telomere shortening during each DNA replication event becomes an irreversibly changing reference point that allows cells to sequentially express a diversity of developmentally regulated genes, which likely emerged in different cells via duplication and mutation (Di Giacomo et al., in preparation). This simple constellation of intracellular activity, which is basic to all metazoan cells, surely represents a primitive form of embryogenesis. Further, the absence in bacteria and unicellular eukaryotes of this simple yet sophisticated mechanism may in part explain why their gene number did not increase as it did in metazoans: i.e., the number of sequences that could be expressed in coordination in the same cell had likely reached its maximum. Gene number could increase only if new sequences (e.g., Hox genes) could be expressed not only in the same cell but also differentially in different cells as cell number increased. Under these conditions, the likelihood of maintaining new genes would increase, as would the potential for generating, exploiting, and/or recruiting genes with new functions (e.g., Hox genes and DNA sequences belonging to Gene Regulatory Networks, GRNs).

The reiterative process outlined above would have persisted because it made possible the maintenance and expression in developmental time and space of new genes that would allow embryos to become increasingly complex over time (i.e., in hindsight, to "produce" evolution). Since the frequency of duplication events is high (Lynch and Force 2000), if the probability of maintaining and preserving these events increased, so too would the total amount of DNA. This hypothesis makes sense of the fact that prokaryotic and lower eukaryotic DNA is Mbp long while metazoan DNA is up to Gbp long.

Recently, the international project ENCODE demonstrated that 80.4 % of the human genome contains elements that participate in at least one biochemical RNA- and/or chromatin-associated event in at least one cell type. This, of course, contradicts the long-held belief that the human genome comprises mostly "junk" DNA. ENCODE also

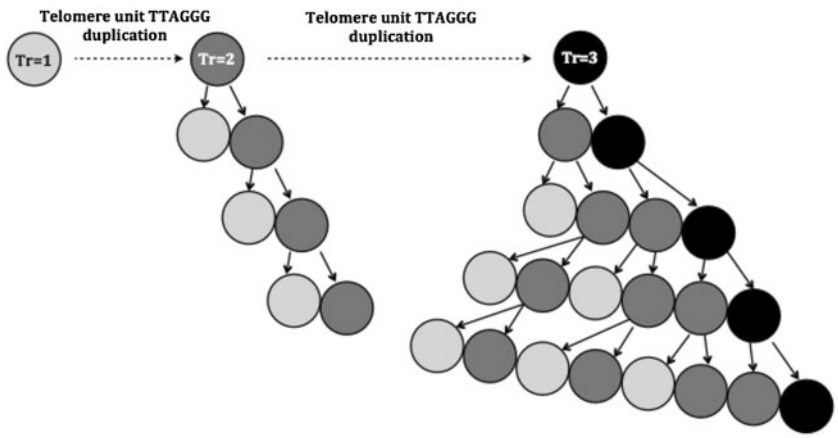

Fig. 1 Asymmetric telomere shortening and cell proliferation in a telomerase-negative background. *Tr1* Telomere repetition is 1 (TTAGGG). *Tr2* Telomere has been duplicated once (TTAGGGTTAGGG). *Tr3* Telomere has been duplicated twice (TTAGGGTTAGGGTTAGGG). *Light gray circles* represent cells with a single telomere repetition (TTAGGG); they stop dividing and enter apoptosis. *Gray circles* represent cells in which one telomere duplication event occurred (TTAGGGTTAGGG); due to the discontinuous synthesis of the DNA, both strands are fully copied, and each gives rise to a shortened (*light gray*) and a full-length (*gray*) telomere daughter cell. A cell whose telomeres have twice duplicated (TTAGGGTTAGGGTTAGGG) (*black circle*) generates a *gray* and a *black* cell. *Gray* (telomere-shortened) cells possess two telomeric repetitions, while *black cells* maintain full-length telomeres (TTAGGGTTAGGGTTAGGG). The repetitive process occurring at each cell division is a step of recursion

showed generally for metazoans that inter-gene spaces contain enhancers and promoters, as well as numerous regions that encode RNA transcripts that are not translated into proteins but which appear to play a role in gene regulation (Ecker 2012). Further, more than 50 % of sequenced metazoan genomes consist of repetitive DNA. Consequently we hypothesize that the repetitive portion of the telomeric and sub-telomeric regions of metazoan DNA serves as a multidimensional reference point that enables different patterns of gene expression such that an embryo can emerge if cell, cell-lineage and, thus, organismal death are delayed.

Differences Between Prokaryotic and Metazoan Genomes

There is general agreement that prokaryotic life (e.g., stromatolites) emerged at least 3.7–3.5 gya (Schopf 2000). Morphologically, these fossils are similar to extant bacteria. Eukaryotes appear in the fossil record ca. 2.7 gya. Fossil unicellular eukaryotes with a defined cytoskeleton date to ca. 1.7–1.5 gya (Roger and Hug 2006). True multicellular and spatially organized organisms are approximately 550 mya (Cambrian explosion; Koonin 2007). Since countless genetic and molecular studies indicate eukaryote monophyly, it is reasonable to conclude that this group descended from a DNA-based prokaryote. Fossils representing most extant metazoan phyla appear in a very short time during the Cambrian (possibly 10–20 mya), thereby suggesting a "big bang" origin for these clades (Cummings 2006).

Although prokaryotic, eukaryotic, and metazoan cells share basic processes (e.g., DNA structure, genetic code, protein synthesis, DNA replication, RNA transcription machineries, and cis- and trans-modes of regulating gene expression), they differ at the level of gene structure, genome architecture, and mRNA processing (Lynch 2006). Lower eukaryotes underwent a moderate increase in gene number, while metazoans experienced significant genomic expansion.

Eukaryotes possess a defined nuclear membrane and linear chromosomes rather than the single, circular molecule of prokaryotes. Eukaryotic cells also comprise specialized organelles, such as mitochondria in metazoans (and chloroplasts in plant cells), that contain a small circular DNA that codes for a few mitochondrial (and chloroplast) genes. Eukaryotic cells are diploid and undergo meiosis as well as mitosis.

Prokaryotic genomes are essentially organized into operons composed of sets of (often functionally related) genes that are co-transcribed in long polycistronic mRNAs. In lower eukaryotes and most metazoans, a gene is regulated by its own promoter. Metazoan genes (but to some extent also those of lower eukaryotes and some Archaea) are subdivided into coding (exons) and non-coding sequences (introns). A long mRNA, containing both introns and exons, is first transcribed into a heterogeneous nuclear RNA that is then processed into a mature mRNA that is much shorter than its corresponding DNA sequence (gene). Gene splicing likely permitted the origination of modular proteins via exon shuffling (Liu and Grigoriev 2004; Schmidt and Davies 2007). Lower eukaryotic and archaean genes may contain short introns, but metazoan genes have large, generously interspersed introns that can be longer than exons. In eukaryotic cells, introns are selectively spliced by spliceosome, which is a complex molecular machine. In archaean cells endoribonuclease cuts exon–intron junctions (Lykke-Andersen et al. 1997). Eukaryotic nuclear genomes also comprise two distinct classes of mobile genetic elements: cut-and-paste transposons and copy-and-paste retrotransposons, which can move to different locations in the genome as it continually reorganizes. For example, the human genome contains ca. 100 mobile genetic elements for each protein-coding gene (Lynch 2007).

With a density $\leq$98 % for protein-coding DNA, prokaryotic genes are tightly packed in their genomes. Although prokaryotes are highly sophisticated "biological machines," their mechanism of regulation is simple, with elementary regulatory circuits (cis-regulation) formed by short nucleotide sequences that, in association with a low number of trans-acting proteins, are responsible for gene expression. Such circuits allow microrganisms to adapt to environmental change because they can reversibly switch on and off specific genes. Thus, most of a prokaryotic nucleotide sequence is read through the genetic code in nucleotide triplets (codons) while the remaining ca. 2 % reversibly regulates the expression of the downstream coding sequences. This ratio of coding to non-coding is reversed in metazoans. In mammals (including *Homo sapiens*), ca. 3 % of total DNA encodes proteins.

Recently, through ultra-deep sequencing of RNAs, ENCODE discovered that ca. 75 % of the human genome is at some point transcribed in some cells. Although these RNAs (shorter than 200 nucleotides) are not translated into proteins, they have regulatory roles (Bernstein et al. 2012). The smaller proportion of the genome consists of untranslated regulatory sequences (promoters, enhancers, silencers, etc.), mobile genetic elements, and extensive repetitive sequences, etc. (Lynch 2007). The change from a pro- to eukaryotic condition in amount of coding versus non-coding DNA occurred in plants and animals either in spite or because of an increase in total DNA. Regardless, total DNA increased from lower eukaryotes, leveling off in flowering plants with 10,000 times as much DNA but only

100 times the number of coded proteins. Further, while in prokaryotes maximum gene density approaches 1,000 genes/Mb, in higher eukaryotes gene density is only ca. 12 genes/Mb. Analyses of entire genomes show a "trend" from prokaryotes to lower eukaryotes to metazoans in increase in gene number, genetic mobile elements, intron number and size, and DNA, repetitive sequences, as well as in genome size, dimensions of intergenic-spacers, and complexity of regulatory regions.

A unique mechanism exploited by metazoans is alternative splicing (Nilsen and Graveley 2010), whereby a heterogeneous nuclear RNA (hnRNA) can be processed with the result that all exons are not represented in the mature mRNA. Consequently, one hnRNA codes for more than one protein: e.g., in the human genome almost 95 % of identified genes incorporate as many as 100,000 splicing sites (Nilsen and Graveley 2010). The emergence of this modality of information retrieval enabled cells to increase protein number without increasing the amount of DNA dedicated to coding regions or to proteins (or RNA) necessary for gene expression. The capacity to generate numerous proteins from the same coding sequence makes understandable the apparent paradox of simple organisms such as *D. melanogaster* and *C. elegans* having 14,000 and 19,000 genes respectively, while complex organisms such as humans possess only ca. 22,500. An extreme case of alternative splicing is noted in the *Drosophila Dscam* gene (Down syndrome cell adhesion molecule), which can assemble 38,016 different mature mRNAs from a single gene (Schmucker et al. 2000).

Although DNA polymerase is necessary for DNA replication in pro- and eukaryotes, what matters is that prokaryotes possess a circular DNA molecule while eukaryotic DNA is distributed in linear chromosomes capped by telomeres. Prokaryotic DNA polymerase can copy the entire DNA molecule at cell division because DNA synthesis proceeds in a 5′⇒3′ direction on both strands. However, eukaryotic DNA polymerase can copy fully only the leading strand because it requires a primer (Okazaki fragment) and a preexisting 3′-OH that allow the addition of free nucleotides to the 3′ end of the newly forming strand; these added nucleotides do not pre-exist in chromosome termini and cannot be synthesized by DNA polymerase. Since eukaryotic DNA polymerase cannot copy the terminal end of the lagging strand (= discontinuous synthesis of the lagging strand) because it lacks the 3′-OH, the terminal region will shorten at each cell division. One "solution" to chromosome shortening and consequent cell or cell-lineage death is to forestall the inevitable by increasing chromosome length. The ancestral eukaryote apparently "did" this by adding short DNA fragments (telomeres) to the ends of chromosomes.

Telomeres consist of short, highly conserved repeats of six nucleotides, usually TTAGGG. Although differing in length, telomeres occur in lower eukaryotic and all metazoan cells. Although telomeres do not code for proteins or RNA products, they maintain chromosome stability (Jain and Cooper 2010) because they compensate for the semiconservative replication of DNA termini (Fig. 2). Telomeres are synthesized by the enzyme telomerase, which is present only in eukaryotic cells. Telomere length varies from ca. 300–600 bp in yeasts, to ca. 15 kb in humans (Jain and Cooper 2010), ≥150 kb in the mouse (Louis and Vershimin 2005; Lynch 2007), and 160 kb in the plant *Nicotiana tabacum* (Fajkus et al. 1995). Lynch (2007) suggests that increased telomere length facilitated the emergence of multicellularity.

Bacterial genomes contain short repetitive sequences—including transposable elements—that while being somewhat dispersed tend to be localized in intergenic regions; they play a role in DNA bending and recombination. These elements are not numerous, and if a DNA fragment containing one duplicates and then recombines, the extra copy is excised. In contrast, repetitive sequences constitute a large percentage of higher eukaryotic DNA. Some of these repetitive DNA elements are transposons, retrotransposons, and pseudogenes, which may consist of a few or even tens of thousands of copies. Copies of other highly repeated DNA sequences (e.g., simple sequences and satellite DNA) may number in the hundreds of thousands or even millions. Repetitive sequences represent more than 50 % of the noncoding region of the human (and generally mammalian) genome. Interestingly, abnormal expansion of repetitive DNA is associated with numerous human genetic defects (Cummings and Zoghbi 2000; Martínez and Blasco 2011).

Although lower eukaryotes differ from prokaryotes in genome organization (e.g., drastic reduction in number or absence of operons, presence of linear chromosomes, introns/exons, increase in amount of DNA), they are not that disparate in gene number: i.e., gene number in prokaryotes ranges between 2,000 and 6,000, with an average of 5,000 genes per microorganism, while in lower

Fig. 2 Schematic representation of a eukaryotic chromosomal end. A telomere consists of long (TTAGGG) repeats. Sub-telomeres are segments of DNA containing repetitive stretches of DNA. The *number* of Hox genes in a cluster reflects the complexity of a specific organism

eukaryotes it ranges between 2,000 and 11,000 with an average of 6,000 genes per organism. Such relative stability in gene number over at least 3.7 gya is interesting in light of the increasingly documented phenomenon of horizontal gene transfer in prokaryotes.

All metazoans possess similar mechanisms of gene regulation, genome organization, basic genome architecture, and G protein-coupled receptors (GPCRs) (signaling proteins). Metazoans are further distinguished in having cell-adhesion proteins (King 2004) and genes that contribute to the formation at least in the larval state of symmetrical body plans (e.g., Hox genes; Mallo et al. 2010).

Prokaryotic and metazoan genes differ in organization and mechanism of regulation. A prokaryotic gene is a simple information unit composed of an uninterrupted nucleotide sequence that is transcribed into an mRNA that codes for a corresponding amino acid sequence (a given coded protein) by the process of translation. Prokaryotic (and eukaryotic) gene expression is controlled by trans-acting regulatory proteins that in prokaryotes are identified as repressors or activators and, in eukaryotes, as transcription factors (TFs). These proteins recognize and bind to specific nucleotide sequences (cis elements). Some genes are constitutively expressed (i.e., transcribed under any circumstance), while others are transcriptionally (conditionally) regulated by reversibly activated regulatory proteins. A prokaryotic cell contains ca. 100 different TFs. In contrast to the yeast genome with about 300 TFs, the *Drosophila* genome contains more than 1,000, while the human genome has $\geq$1,800 (Bernstein et al. 2012). It thus seems reasonable to suggest that genomic diversity and organismal complexity probably emerged coincident with more elaborate mechanisms of regulating gene expression.

Prokaryotic genes can be reversibly and transcriptionally activated by changes in the environment (e.g., presence/absence of low molecular weight molecules) that regulate the activity of constitutive regulatory factors necessary to transcribe genes downstream to a controlling operon. These regulatory proteins bind upstream to short cis elements and downstream to the binding site of RNA polymerase that transcribes (or represses) polycistronic mRNAs, which are translated into corresponding proteins. This mode of regulation is widespread among prokaryotes and allows the organism to adapt quickly to environmental change by the reversible binding to activators/repressors of low molecular weight inducers (e.g., a sugar, an amino acid, etc.). Further unlike in eukaryotes, prokaryotic transcription and translation occur simultaneously. These simple circuits are similar to mechanical/electrical switches or simple electrical devices that operate in a reversible fashion (i.e., a first-order Boolean logic machine).

Most bacterial gene products are described as "housekeeping genes" that code for enzymes and structural proteins as well as a small number of regulatory factors that control the syntheses of these proteins. In other words, the information (structural proteins, enzymes) to build a microorganism is coded in the DNA, while regulatory factors allow their synthesis in response to environmental change by extracting the information coded in the nucleotide sequences of genes. A direct relationship between genes–proteins–phenotype thus exists within the same cell (the central dogma of biology, CDB; Crick 1958). By conjugation or transformation, a gene or a set of genes, such as genomic islands, can be integrated into and thus add new functions to a bacterial chromosome. Gain or loss of these genes (new modules) allows bacteria to adapt to new environmental conditions but not to perform more elaborate tasks, such as achieving multicellularity or complex form.

An advantage provided by gene organization based on single-gene transcripts rather than operon-based polycistronic mRNAs is that a gene controlled by its own promoter is independently mutable, while all genes within an operon are affected by mutations controlling transcription initiation. It has been suggested that local duplication of a promoter region and single base substitutions in newly duplicated DNA fragments can induce the formation of more complex enhancer elements (transcription factor binding sites) that can "accept" the binding of new regulatory factors.

Prokaryotic operators (cis-regulatory elements) lie immediately upstream to the transcriptional initiation site and downstream to the RNA polymerase binding site, thus allowing interaction with the corresponding regulatory factors. It is reversed in eukaryotes, in which TFs bind to long cis-regulatory regions located upstream to the RNA polymerase binding sites. In metazoan genes these regions can be >1 Mb compared to short, ca. 100 nt, prokaryotic regions. In addition to long upstream regulatory regions, metazoan genes may include additional regulatory regions downstream to the end of the gene and may even contain regulatory regions within their introns.

For the 147 human cell lines analyzed, ENCODE identified more than 400,000 active enhancers, more than 70,000 regions with promoter-like features, 2.9 million sequences recognized by putative TFs, of which a third are uniquely active in a given cell type, and only 3,700 regulatory sequences, as well as 20,687 protein-coding genes with, on average, 6.3 alternatively spliced transcripts (3.9 different protein-coding transcripts) per locus. In total, GENCODE-annotated exons of protein-coding genes cover 2.94 % of the genome or 1.22 % for protein-coding exons. ENCODE also delineated 11,224 pseudogenes, of which 863 were transcribed and associated with active chromatin. Taken together, this indicates that transcriptional differences between different cell types are enormous (Bernstein et al. 2012).

Metazoan Development

Multicellular organismal development requires expression of a finely tuned genetic program. Hox genes code for TFs that determine pattern formation along anteroposterior and lateral embryonic axes. Hox genes occur in clusters toward one of the chromosomal telomeric regions and their relative positions reflect the timing of their expression. In embryos of genetically modified mice, Hox genes positioned nearer the telomeric extremity are expressed earlier than those closer to the centromere (Fig. 2) (Tschopp et al. 2009). Consequently, GRNs are dependent on the timing of Hox gene expression and diffusible TFs are not operative if Hox genes are not properly ordered and in the correct physical location.

Telomerase Down-regulation

After an egg is fertilized telomerase is only active for a species-specific number of cell divisions, after which it is down-regulated (Taylor and Delany 2000), with the result that embryonic-cell telomeres shorten at each cell division. Because telomerase is down-regulated in metazoan somatic cells, they undergo a fixed and limited number of cell divisions (the Hayflick limit). Upon reaching this limit, chromosomes shorten because DNA polymerase cannot replicate a chromosome end; eventually, cells die. Since telomerase plays a key role in maintaining chromosome length, it is puzzling why in metazoans (as well as in plants) genes coding for telomerase (both TERT and TERC components) are down-regulated after the initial stages of embryogenesis.

This question is critical because repression of telomerase activity in metazoans results not only in telomere shortening (Liu et al. 2007) but also in organismal senescence and eventual death. In humans, alterations of the repetitive sequence of the subtelomeric region cause various teratological disorders (Heutink et al. 1994). Conversely, reactivation of telomerase produces cancer cells, whose descendants essentially become "immortal" lineages since they do not experience telomere erosion. Further, in most advanced cancers, telomerase has the capacity to directly regulate cancer-promoting pathways (Artandi and DePinho 2010).

Genomes and Computer Analogies

Metazoan genomes execute a developmental genetic program that, from a fertilized egg onward, produces hundreds of differently specialized cells that perform the myriad tasks necessary to determine an organism's body plan. The fertilized egg's cytoplasm contains the "startup information" (like a computer "bootstrap") that provides the instructions crucial to initiating a program encoded in the DNA. Afterward and without external input, this program is capable of reading more program instructions and becoming a self-sustaining recursive process that initiates cell differentiation (i.e., "runs" the program) (Gilbert 2010).

A non-nuclear bootstrap is required because start-up instructions (an external "basic routine") are and cannot be part of the program encoded in the nuclear DNA (metazoan genome at t_0). In other words, nuclear DNA at t_0 cannot begin to unravel its encoded instructions unless the initial conditions for transcribing "the first gene" are provided by a set of proteins (or mRNAs) not coded in the genome at t_0. In order for the genome to begin unfolding its interconnected regulatory circuits, it must express "a first gene," which requires at least one protein (a TF) to instruct RNA polymerase to transcribe that gene. However, that TF demands a gene coding for it, and so on, ad infinitum. This, of course, is biologically unsustainable, but makes understandable why a fertilized egg's nuclear DNA must obtain its "boot up" information from a source external to it in order to transcribe the "first gene" that then initiates the developmental program.

This mode of operation is equivalent to a computer. When one launches a computer program, the program does not contain the information necessary to start itself, because a computer program requires a bootstrap that sets the conditions to launch that program when the computer system is turned on. Further, a computer system needs an external set of instructions, also located in the booting system, in order to shut down itself and the program. In analogy, when a developmental genetic program "ends," it cannot initiate the system in the next generation of offspring without the proteins and mRNAs in the egg's cytoplasm (generally described as the "maternal effect"). But once this initiation occurs, the unfolding genetic program induces differential gene expression of sets of instructions that then produce cells expressing different groups of genes. This, in turn, generates an asymmetry in the capacity to transcribe specific genes in differently developing cell masses.

This process of extracting genomic information leads to the differential expression of genes in time and in space within the emerging embryo (Wyrick and Young 2002; Davidson 2006). That is, after "booting up," different combinations of existing TFs operate in different cells or, as in *Drosophila* embryos, in segments that initiate the regulation of differential gene expression in various domains by binding to specific cis-regulatory regions (Ben-Tabou de-Leon and Davidson 2007). This establishes specific modular networks of genes that will eventually generate sub-circuits with specific, localized functions. Such cis-regulatory modules comprise different binding

sites for multiple TFs that activate specific genes in specific regions of the embryo (Davidson 2010).

Unlike in prokaryotes, the unfolding metazoan genome is organized into numerous cis-regulatory modules and TFs that control the highly sophisticated gene networks that lead from a single cell to the formation of a multicellular and topologically differentiated organism. These complex sets of instructions have the capacity to generate different groups of specialized cells within the embryo as well as "dynamic topological objects" (groups of cells) that undergo continuous deformation and modification via ever-changing instructions from specific genes and local circuits. Consequently, defined but acting independently, regulatory networks are capable of generating distinct forms and biological properties that can eventually explore different topological landscapes.

The Turing Machine and Development

A central problem in biology is deducing the characteristics of animals solely from their nucleotide sequences. Studies in molecular biology and development have shown that multicellular organisms possess several classes of genes with distinct hierarchical roles. Master regulator genes (e.g., Hox genes), which establish an animal's body plan, are at the highest level. At the lowest level are genes that encode metabolic or structural proteins—the materials of which cells are made—necessary for building a cell and maintaining energy flow. Between these levels are other gene classes whose product interactions establish a complex network. These GRNs (Davidson 2010) have been extensively studied in several model systems, such as *C. elegans*, the sea urchin, *Drosophila*, and the mouse, and described as logic circuits similar to computer Boolean logic gates (Davidson 2006; Shapiro 2006).

Computers are based on the Turing machine, which is a theoretical device with two elements: (1) a table of rules that manipulates symbols on a tape, and (2) the tape itself, on which symbols are written. A sequence of symbols is an algorithm. The table of rules is a consistent and complete system that can be implemented by logic gate networks. We suggest that the GRNs described by Davidson and others represent the tables of rules of a Turing machine. Cells therefore possess the table of rules but not the tape. Consequently, in order to equate "a cell" with "a computer," there must exist a "tape" that allows a cell to function as a Turing machine. In other words, during embryogenesis, at a certain time, t_1, a given number of identical cells, n_1, possess a logic network that represents the functioning interconnectivity of different gene products. Input into the network at time t_1 derives from the previous network status at time t_0, with $n_0 < n_1$. In turn, output products of the network at time t_1 will feed the network at time t_2, with $n_2 > n_1$. By means of successive cell divisions controlled not by the circuits, but by the cell cycle, an embryo proceeds from time t_0 to t_1 to t_2 with cell number $n_0 < n_1 < n_2$.

The question this raises is: what controls the timing of cell division and an increase in cell number from n_0 to n_1 to n_2? This point is crucial because, when a new logic circuit is established from a previous one, it is tantamount to adding a new Boolean logic network to the circuit. Yet what has so far been explored in the literature on the regulation of development is the functional tree of an algorithm, which runs in successive "frozen states." In this scenario, the entire circuits at t_0, t_1, and t_2, etc., constitute the table of rules of a Turing machine. But, in order to be complete, a Turing machine also requires a tape that, in the case of development and differentiation, moves the system from cell number n_0 to n_1 to n_2 and so forth. Thus, every time a gene high up in the hierarchy participates in a circuit, it must first read from the tape in a recursive manner steps 0, 1, 2, etc. In a real computer, the tape encompasses central memory, a clock, and a counter. In a computer that performs only one program, the central memory resides in the CPU and only a clock and counter are required. In order for a cell to act as a tape, it must have an irreversible counter that is immune to manipulation by the system itself.

We propose that telomere shortening, which results from metazoan cells' inability at each cell division to copy fully the ends of their chromosomes, represents a biological irreversible counter. In this regard it is noteworthy that telomere length and cell cycle control are linked (Marcand et al. 2000) and regulated by some of the same proteins (e.g., TRF1 in mammals) that bind the double-stranded telomeric TTAGGG repeats into the t-loop that caps and protects chromosome ends from erosion and aberration (Dubrana et al. 2001). Further, conceiving metazoan cells as Turing machines acknowledges their capacity to define the recursive functions necessary to express topology, i.e., complex form. Since lower eukaryotes and bacteria do not have a tape, they do not experience telomere erosion and, according to our model, could never achieve multicellularity.

Lastly, it has long been known that if a heat stress (or membrane perturbing agent, e.g., ether) is applied to an embryo at certain times, stress proteins do not accumulate to normal levels and thus cannot fully protect cellular functions (Petersen and Mitchell 1987). Under such stress, an adult develops an altered phenotype (phenocopy), even though mutation did not occur. Consequently, a cross between two phenocopy individuals produces normal offspring. In the case of a *Drosophila* phenocopy, the altered fly has a duplicated abdomen (bithorax) with four wings (two per thorax) and is phenotypically similar to

Drosophila in which thorax duplication was caused by Hox gene mutation (Ho et al. 1983; for discussion of phenocopy in an evolutionary perspective see Maresca and Schwartz 2006). Our model predicts that the bithorax phenocopy will emerge upon temporary cessation of telomere erosion due to stress/perturbation that alters the folding and functionality of proteins, including those involved in regulating gene expression (e.g., telomere erosion) and duplication of cells responsible for thoracic development. With subsequent resumption of telomere shortening, these two groups of cells share the same external reference point, undergo the same number of divisions, activate the same circuits, and yield a double thorax.

Prokaryotic and Metazoan Genome-Architecture Information and Gödel's Theorems of Incompleteness

The most significant difference between single-cell and multicellular organisms, especially metazoans, is that the latter comprise simultaneously different kinds of cells that, with the same DNA sequence, perform different, recursive, homogeneous, and tightly orchestrated developmental programs. Further, multicellular organisms, such as plants, and various metazoan organs, such as kidneys and lungs, express typically self-similar patterns (fractals), as well as other recursive structures. Since fractal geometry is a heritable property of biological organisms, and fractality is a process of recursion, it must be coded in their genomes.

Development is the topological transformation of an increasing mass of cells that begins as a single cell (the zygote) and culminates in an adult organism with a specific shape. Since topological forms, such as those generated during embryogenesis, are described mathematically with recursive functions, we hypothesize that plants and animals could only achieve shape if they embodied a recursive function that was implemented by an external counter. We suggest further that telomere shortening is the counter (index) for the "recursive program" of embryogenesis and that this counter differentially drives the sequential expression of Hox genes and the mobilization of GRNs. When telomeres become entirely eroded, the organism enters senescence and dies.

It is well known that if a system is formalized such that it defines a recursive process, it is equivalent to arithmetic. We can therefore use the terms "recursion" and "arithmetic" interchangeably. A prokaryotic or a lower eukaryotic cell cannot run a recursive developmental program because neither the architecture of the genome nor the mode of processing information is self-referent. Among the different levels of potential inquiry, we are here concerned with the higher levels of information processes (IPs) that act above but not within the lowermost layer of the stack, the CDB, which we conceive as basic hardware. The functionality of prokaryotic, lower eukaryotic, and metazoan cells can be represented by a conceptual model, in which stacks of different layers subdivide the system into smaller components (Fig. 3).

The Transcription Regulatory System (TRS) of prokaryotic cells is a simple, inducible mechanism of gene regulation. From an IP perspective, the TRS of a prokaryotic cell can be formalized as binary switches that are

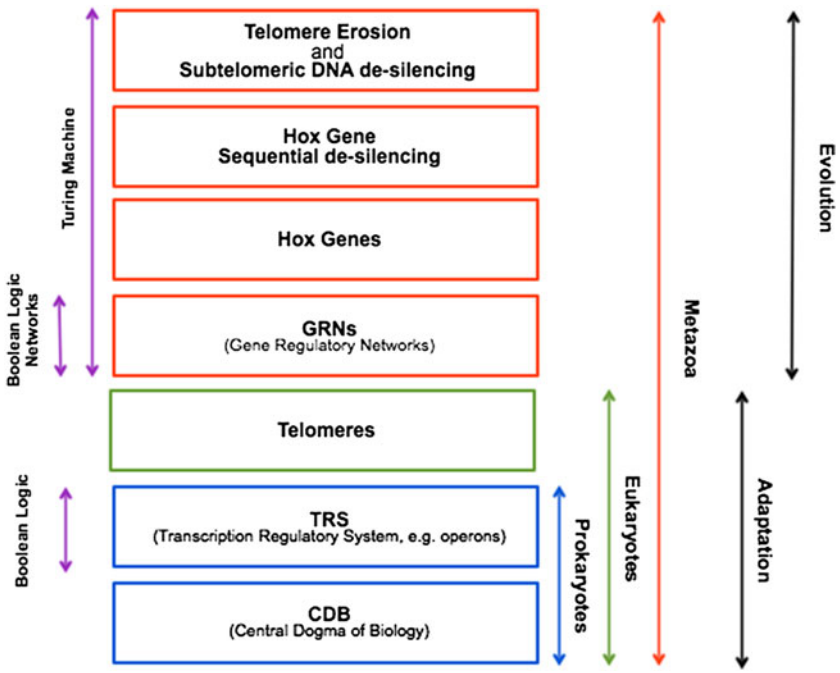

Fig. 3 Stack organization of prokaryotic, eukaryotic, and metazoan genomes. Prokaryotic cells are capable only of Boolean logic and lower eukaryotic cells only of Boolean logic networks. Only the metazoan genome can implement a Turing machine and thus produce complex forms (topology)

part of Boolean algebra. Since Boolean algebra is self-coherent and complete, so, too, must be the IP-layered formal system of a prokaryotic cell. In contrast, the multiple IP layering of metazoan cells requires formalization of a complex system that leads to recursion and to Gödel's theorems of incompleteness, which have proven applicable to diverse disciplines (Hofstadter 1979).

Gödel's theorems rest on two assertions: (1) if a formal system capable of expressing arithmetic is consistent, it cannot be complete, and (2) the consistency of the axioms cannot be proven within the system itself (Nagel and Newman 2001). To quote Kleene's (2002, p. 250) representation of Gödel's first theorem of incompleteness: "Any effectively generated theory capable of expressing elementary arithmetic (here = Metazoa) cannot be both consistent and complete. In particular, for any consistent and effectively generated formal theory that proves certain basic arithmetical truths, there is an arithmetical statement that is true, but not provable (here = telomerase)." Therefore, a true metazoan cell state and cell-lineage immortality cannot coexist.

Conclusions

In summary, we have argued the following:

Telomere shortening represents a crucial external reference point that makes possible the development of complex topology because it permits recursion.

Fractals in nature (e.g., coast, cloud, or mountain contours) result from physical laws and can be described a posteriori by fractal mathematics. In contrast, fractal geometry in metazoans (e.g., lungs alveoli, kidney, etc.) and plants (e.g., shapes of trees, leaf distribution) appears to emerge from within the organism, which would necessitate a capacity for computing similar to computer software. If true, fractals would emerge in biological organisms if and only if they embody a "machine" capable of computation.

The change from a circular, nucleotide-sequence dependent genomic organization to a complex, linearly separate chromosomal organization characterized by down-regulation of telomerase and concomitant telomere shortening was essential for the emergence of multicellularity and complex topology.

Lack of telomerase activity is essential for cell development, but the trade-off for organismal differentiation and complexity is eventual cell-lineage and organismal death.

Acknowledgments We wish to dedicate this paper to the late Leonardo Coen Cagli with whom we shared the passion for numbers and mathematics.

Appendix

Definitions

Fractal A fractal is an object or quantity that displays self-similarity on all scales (Wolfram 1999, p. 986). Mathematically rigorous treatment of fractals can be traced to functions studied in the 19th century by Karl Weierstrass, Georg Cantor, and Felix Hausdorff. In 1975, as a result of analyzing continuous but not differentiable functions, Mandelbrot (1977) coined the term "fractal." A mathematical fractal is based on an equation that undergoes iteration, which is a version of feedback that is based on recursion. In nature, fractals such as those that constitute mountain and coast contours, tree shape, lung alveoli, etc., obey physical laws. Other examples of fractals are defined as portraying exact self-similarity, quasi self-similarity, or statistical self-similarity.

Gödel's Theorems of Incompleteness Gödel's (1931) two incompleteness theorems are theorems of mathematical logic that establish the inherent limitations of all but the most trivial axiomatic systems capable of doing arithmetic. These theorems are important in mathematical logic as well as in the philosophy of mathematics because they prove impossible Hilbert's program to find a complete and consistent set of axioms for all of mathematics.

Gödel's first incompleteness theorem states that no consistent system of axioms, whose theorems can be listed (i.e., generated from the axioms) by an "effective procedure" (automatic reasoning), is capable of proving all truths about the relations of natural numbers (arithmetic) (Nagel and Newman 2001). For such a system, there will always be statements about natural numbers that are true but cannot be proven within the system; in other words, a true statement cannot be proven using the rules of the system. Thus no formal system that satisfies the hypotheses of the theorem and also strives to characterize natural numbers can actually do so because there will always be true number-theoretical statements the system cannot prove.

Gödel's second incompleteness theorem is a corollary of the first. It demonstrates that such a system cannot also validate its own consistency. For any formal, effectively generated theory, T, including basic arithmetical truths and certain truths about formal provability, T includes a statement of its own consistency if and only if T is inconsistent. This theorem strengthens the first because the statement constructed does not directly express the consistency of the theory. The proof of the second incompleteness theorem is essentially obtained by formalizing the proof of the first incompleteness theorem within the theory itself.

Turing Machine The "Turing" machine was described in 1936 by Alan Turing, who characterized it as an "automatic-machine." It was not intended as a practical computing technology, but as a thought experiment that represented a computing machine. Turing machines help computer scientists understand the limits of mechanical computation. A Turing machine is therefore a theoretical device that manipulates symbols on a tape according to a table of rules. In spite of its simplicity, a Turing machine can be adapted to simulate the logic of any computer algorithm and is particularly useful in explaining the functions of a CPU inside a computer.

In 1948 Turing wrote that the Turing machine, which he then referred to as a Logical Computing Machine, consists of "an infinite memory capacity obtained in the form of an infinite tape marked out into squares, on each of which a symbol could be printed. At any moment there is one symbol in the machine; it is called the scanned symbol." Although the machine can alter the scanned symbol and its behavior is determined in part by that symbol, the symbols elsewhere on the tape do not affect the behavior of the machine. However, one of the elementary operations of the machine is that the tape can be moved back and forth through it.

Recursion Recursion is a process of repeating items in a self-similar manner. For instance, when the surfaces of two mirrors are parallel, the resultant nested images represent a form of infinite recursion. The term has different meanings in disciplines ranging from linguistics to logic. The most common application of recursion is in mathematics and computer science, in which the term refers to a method of defining functions in which the function being defined is applied within its own definition. Specifically, this defines an infinite number of instances (function values), using a finite expression that, for some instances, may refer to other instances, but in such a way that no loop or infinite chain of references can occur. The term is also used more generally to describe a process of repeating objects in a self-similar way.

References

Artandi SE, DePinho RA (2010) Telomeres and telomerase in cancer. Carcinogenesis 31:9–18

Ben-Tabou de-Leon S, Davidson EH (2007) Gene regulation: gene control network in development. Annu Rev Biophys Biomol Struct 36:191–212

Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, ENCODE Project Consortium et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74

Brusca RC, Brusca GJ (2003) Invertebrates, 2nd edn. Sinauer Associates, Sunderland

Crick FHC (1958) The biological replication of macromolecules. Symp Soc Exp Biol 12:139–163

Cummings FW (2006) On the origin of pattern and form in early Metazoans. Int J Dev Biol 50:193–208

Cummings CJ, Zoghbi HY (2000) Trinucleotide repeats: mechanisms and pathophysiology. Annu Rev Genomics Hum Genet 1:281–328

Davidson EH (2006) The regulatory genome: gene regulatory networks in development and evolution. Elsevier, Amsterdam

Davidson EH (2010) Emerging properties of animal gene regulatory networks. Nature 468:911–920

Dubrana K, Perrod S, Gasser SM (2001) Turning telomeres off and on. Curr Opin Cell Biol 13:281–289

Ecker JR (2012) ENCODE explained. Nature 489:52–53

Fajkus J, Kovarik A, Kralovics R, Bezdek M (1995) Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. Mol Gen Genet 247:633–638

Gilbert SF (2010) Developmental biology, 9th edn. Sinauer, Sunderland

Gödel K (1931) Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme, I. Monatshefte für Mathematik und Physik 38:173–198

Heutink P, Zguricas J, van Oosterhout L, Breedveld GJ, Testers L, Sandkuij LA, Snijders PJ, Weissenbach J, Lindhout D, Hovius SER, Oostra BA (1994) The gene for triphalangeal thumb maps to the subtelomeric region of chromosome 7q. Nat Genet 6:287–292

Ho MW, Bolton E, Saunders PT (1983) Bithorax phenocopy and pattern formation. I. Spatiotemporal characteristics of the phenocopy response. Exp Cell Biol 51:282–290

Hofstadter D (1979) Gödel, Escher, Bach: an eternal golden braid. Vintage Books, New York

Jain D, Cooper JP (2010) Telomeric strategies: means to an end. Annu Rev Genet 44:243–269

King N (2004) The unicellular ancestry of animal development. Dev Cell 7:313–325

Kleene SC (2002) Mathematical logic. Orig. 1967. Dover, New York

Koonin EV (2007) The biological big bang model for the major transitions in evolution. Biol Direct 2:1–17

Koziol C, Borojevic R, Steffen R, Müller WEG (1998) Sponges (Porifera) model systems to study the shift from immortal to senescent somatic cells: the telomerase activity in somatic cells. Mech Ageing Dev 100:107–120

Kruse M, Batel R, Steffer S, Schröder HC, Müller IM, Müller WEG (2000) Sponge homologue to human and yeast gene encoding the longevity assurance polypeptide: differential expression in telomerase-positive and telomerase-negative cells of *Suberites domuncula*. Mech Ageing Dev 118:115–127

Liu M, Grigoriev A (2004) Protein domains correlate strongly with exons in multiple eukaryotic genomes: evidence of exon shuffling? Trends Genet 20:399–403

Liu L, Bailey SM, Okuka M, Muñoz P, Li C, Zhou L, Wu C, Czerwiec E, Sandler L, Seyfang A, Blasco M, Keefe DL (2007) Telomere lengthening early in development. Nature Cell Biol 9:1436–1441

Louis EJ, Vershimin AV (2005) Chromosome ends: different sequences may provide conserved functions. BioEssays 27:685–697

Lykke-Andersen J, Aagaard C, Semionenkov M, Garrett RA (1997) Archaeal introns: splicing, intercellular mobility and evolution. Trends Biochem Sci 22:326–331

Lynch M (2006) The origin of eukaryotic gene structure. Mol Biol Evol 23:450–468

Lynch M (2007) The origins of genome architecture. Sinauer, Sunderland

Lynch M, Force A (2000) The probability of duplicate-gene preservation by subfractionalization. Genetics 154:459–473

Mallo M, Wellik DM, Deschamps J (2010) Hox genes and regional patterning of the vertebrate body plan. Dev Biol 344:7–15

Mandelbrot BB (1977) Fractals: form, chance, and dimension. Freeman, San Francisco
Marcand S, Brevet V, Mann C, Gilson E (2000) Cell cycle restriction of telomere elongation. Curr Biol 10:487–490
Maresca B, Schwartz J (2006) Sudden origin: stress protein concentration and rapid environmental changes underlie a general mechanism of evolution. Anat Rec B New Anat 289B:38–46
Martínez P, Blasco MA (2011) Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. Nat Rev Cancer 11:161–176
Nagel E, Newman J (2001) Gödel's proof. New York University Press, New York
Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463:457–463
Petersen NS, Mitchell HK (1987) The induction of a multiple wing hair phenocopy by heat shock in mutant heterozygotes. Dev Biol 121:335–341
Roger AJ, Hug LA (2006) The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. Philos Trans R Soc Lond B 361:1039–1054
Schmidt EE, Davies CJ (2007) The origins of polypeptide domains. BioEssays 29:262–270
Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101:671–684
Schopf JW (2000) Solution to Darwin's dilemma: discovery of the missing Precambrian record of life. Proc Natl Acad Sci USA 97:6947–6953
Shapiro JA (2006) Genome informatics: the role of DNA in cellular computations. Biol Theory 1:288–301
Taylor HA, Delany ME (2000) Ontogeny of telomerase in chicken: impact of downregulation on pre- and postnatal telomere length in vivo. Dev Growth Differ 42:613–621
Tschopp P, Tarchini B, Spitz F, Zakany J, Duboule D (2009) Uncoupling time and space in the collinear regulation of Hox genes. PLoS Genet 5:1–12
Wolfram S (1999) The mathematica book, version 4. Cambridge University Press, Cambridge
Wyrick JJ, Young RA (2002) Deciphering gene expression regulatory networks. Curr Opin Genet Dev 12:130–136