 doi:10.20944/preprints202210.0371.v1

*Hypothesis*

# Evolution is Governed by the Information Laws of Conservation and Speciation

**Radomir Crkvenjakov**

784 Manor Way, Los Altos, CA 94024, rcrk@comcast.net

**Abstract:** Others have proposed that living systems are characterized by the possession of a kind of information unique to them. This information, variously called semantic, symbolic, or coding information conveys meaning from senders to recipients, most commonly among genome and other cellular components, different cells, organisms, etc., enabling all biological functions. Unlike other kinds of information, it is only indirectly quantifiable and does not increase by copying. Here, the term system information is used to encompass the entire package of coding information enabling all organisms of a species to complete their life cycles through possession of its copy (Crkvenjakov and Heng 2021). System information grows in a nonreductive evolution only during speciation proper by the acquisition of novel functions by daughter species. Contemporary science agrees that species are units of evolution but has failed to recognize and parse the importance of phases of speciation and stasis for evolution.Confinement of system information change to speciation only explains for the first time why there are species and why their number increases nearly exponentially throughout evolution. More importantly, here we define that evolution as a process operating on a planetary scale, while chance based, is non-random mathematically since it obeys the Laws of Information which dictate the behavior of System information in organisms.Growth and persistence of system information on the non-degenerative portion of Tree of Life follow the law of conservation, which states that its growth is as parsimonious (minimal) as possible and the persistence of parental system information the maximal possible in each unit speciation step. This formulation is a restatement of Loewnstain's (1999) principle of information economy and our lock-in law (Crkvenjakov and Drmanac 2007). The principle holds that since, on chance assumption, the system information change cost is directly proportional to time, more costly pathways of achieving the same new functional goal will be, almost without exception, competed out by the less costly ones. As a corollary of the conservation law, the speciation law states that speciation follows the most economical way of growth of system information while conserving the existing one, which is horizontal information transfer by molecular sex (uptake of exogenous nucleic acid into the progeny genome) by exchange among two parents from maximally possible divergent lineages. The speciation law dictates that molecular sex mode be used since it is the most parsimonious one. The molecular sex cost in information bits is zero to the recipient, which is not the case with the generally accepted mode of vertical accumulation of information bits in a population gene pool within a lineage.

**Keywords:** Evolution; Information Laws; Coding Information; System Information; Species; Stasis

## Introduction

Proponents of Intelligent design have attacked Darwinism on the basis that it, when boiled down to essentials, presupposes randomness of mutation occurrences

> *….Richard Dawkins very nicely set out the essentials. Random mutations and natural selection drive biological change. In a still lower gear, random variations and natural selection drive change beyond the margins of an organism. Maddened by the same reproductive urge, bowerbirds construct elaborate nests and men, elaborate cities. If these ideas are now part of the Great Gabble, they remain well below the level of the serious*

 doi:10.20944/preprints202210.0371.v1



> *sciences. They encompass neither the revolution in evolutionary thought introduced by Motoo Kimura's theory of neutral mutations, nor the criticism that an inherently random process cannot produce the complex structures characteristic of molecular biology, let alone human society Environments change, no doubt, for reasons of their own. Those changes must be described by a random variable. The product of two random variables is again a random variable, randomness going straight down to the bottom in Darwinian theories. (Berlinsky 2019)*

The critics interpreted biologists to say that chance in biology operated mathematically just as it does in tossing a coin. That idea has been negated from the very start by biologists, including Darwin himself, Mayr, and more recently, a majority of biologists spearheaded by by EvoDevo practitioners (Shubin 2020). However, this negation was indirect, based on various biological grounds like descent with modification, canalization, evolutionary constraint, and the like, and in molecular biology with the particular evolutionary models for the origin of the specific extant nucleic acids and proteins.

Thus, so far, biologists have not shown a general principle applicable to entire biology, both understandable and acceptable to colleagues from "hard" sciences, why mathematical randomness is excluded from their science. Such general principles would satisfy requirements for biological laws and make biology more intelligible by making it closer to physics and mathematics. We propose two system information laws here as a basis of general theory of biology based on information. The laws do not uncover any new biological reality since they deal with the same evolutionary phenomenon described by Wallace in his "Sarawak Law" of 1854 that new species arise exclusively from the most closely related species (Wallace 1855). However, the laws outlined here give an explanation of causes and consequences of the Sarawak Law for Evolution.

Here we want to show how consideration of molecular system information provides answers to the most basic biological questions; some of which are so near to the core of biology as science that they are never posed and answered in literature. System information has also been called Karyotype code (Heng et al. 2011, Crkvenjakov and Heng 2021) or Macrocode (Prosdocimi and Farias 2021) to emphasize its encoding nature. The phenomena for which causes are sought include exponential reproduction, "molecular sex," diversification, species, individuals, speciation, lineages, genotype, and phenotype and their proper relations. Some of the answers do not challenge the received theory of evolution, but others do, and the paper, on the whole, could be viewed as an argument for the information-based theory of the origin of life and evolution.

Modern Molecular Biology finds itself unable to uphold outdated NeoDarwinism of the Modern Synthesis Theory of Evolution (Shapiro 2011, 2021, Crkvenjakov and Heng 2021). At least three major evolutionary phenomena require novel theoretical explanations, which we find in the analysis of biological information.

First, the expanding evidence of sex among unlikes in the last 50 years has caught evolutionary theory unprepared. From its Aristotelian origins, biology has been based on observation of common events of mating between likes - members of the same species; unlikes, like Leda and the swan, mated only in Greek mythology. Following Darwin's era, this observation of naturalists was codified by the amalgamation of the nascent Mendelian Genetics with Darwin's Theory of Natural Selection in the 1930s in the so-called modern synthesis (MS), which became the textbook evolutionary theory used even today. It is entirely based on vertical inheritance and sex of likes. Contrary contemporary examples are noted but not taken as damaging to the central premises. Sex of both of likes and unlikes constitutes horizontal inheritance with the minority parental contribution varying from less than a single gene to 50% of the genome. Unlikes are defined as cases of mating where parents come from different subspecies, species, classes, or even kingdoms. Since the advent of genomics, previously rare examples of horizontal inheritance have turned into a flood. For example, as of lately, we know that members of our species *H. sapiens.*

 doi:10.20944/preprints202210.0371.v1



mated with members of at least two different species of humans called Neanderthals and Denisovans (Green et al., 2010, Reich et al. 2010); frog species *X. leawis* is tetraploid resulting from fusion of germ cells from the two different diploid frog species 17 million years ago (Session et al. 2016). In the prokaryotic world, horizontal inheritance is so rampant that some have proposed that species do not exist there (Doolittle and Zhaxybayeva 2009, Margulis 2006, Doolittle 2019). If one notes that some demonstrated horizontal inheritance events could have produced new species only in a saltatory manner from the single progeny organism of a single mating, while no such evidence is as clear cut for population speciation modes required for modern synthesis, then the only way for the received theory to survive is to claim the frequency and/or importance of speciations by horizontal inheritance is dwarfed by speciations in the populational mode. We rejected that argument on the basis that clear origin by horizontal inheritance of eukaryotes, vertebrates, teleostean fishes, some amphibian groups and numerous families and species of plants and fungi, and species innovations in prokaryotes is the sufficient cause to propose universality of speciation by horizontal inheritance events (Crkvenjakov and Drmanac 2007).

Second, the theory of evolution based on gradual genetic change did not have use for the commonsense view also shared by embryology and morphology about the existence of unchangeable phenotypic characteristics of any given species. Modern synthesis followed Darwin, explaining these characteristics away as ancestral features that are chiseled away to be eventually lost in the future by action of natural selection on the genome. Evo-Devo results indicated failure of the perpetual change scenario for some important groups of genes (and their phenotypes) whose resistance to change is shown by their evolutionary antiquity. But what about the genome? Is it amenable to change during species' lifetime? In answer, cytological species-specific karyotype pictures have been available for almost 100 years.

Instead of consideration of this view of chromosomes as an easily accessible proxy for a functional genome structure, karyotype was treated by evolutionists as if it depicted an inert scaffold for carrying genes without any particular significance. Consequently, the importance of its change as a wholly speciation-restricted hereditary phenomenon has been either repudiated or entirely overlooked by modern synthesis in opposition to minority views like Goldschmidt's (1940) and others throughout last century. However, since 2000 onward, the population stability of genome organization was demonstrated by the complete genome sequencing of thousands of species and the specificity of 3D organization of chromatin in the interphase nucleus. These results show that mutants with the disturbed genome structure would be inviable, and this is why natural selection is unable to select for substantial change in genome and karyotype organization during species' lifetime. The evolutionary importance of karyotype change was brought back into recent literature by an interesting proposal by Heng (2009, 2019). According to him, the Genome Chaos is the specific evolutionary process of substantial genome reorganization semaphored by the karyotype change that is equally required for successful speciations, as well as for the lethal outcomes of cancer.

Third, another substantial unsolved problem for modern synthesis is the morphological stasis of most species in the fossil record (cf. Jablonsky 2017). Synthesis' proponents advanced explanations based on causes external to organisms, each of which operates today on a much shorter time scale than the near-universal duration of the sameness of organisms for millions of years and generations. Rather than criticize those explanations, we searched for the internal cause of enduring sameness and found it in the widely accepted molecular biology postulate that every trait or groups of interconnected traits in species phenotype, such as zebra stripes or panda's thumb, are the result of action of a specific GRN (gene regulatory network) operating in the species in question. Obviously, the stripe program is an integral part of the overall sum of networks for all traits over the entire life cycle of an organism of a species. We call this integral entity system information or Karyotype Code (Crkvenjakov and Heng 2021) and specify that it is the informational one, materialized through interactions of all species' molecules. What about the fact that all

 doi:10.20944/preprints202210.0371.v1



humans or pandas are not the same? Is not this a contradiction fatal to the above definition? Not if one divides phenotype into two parts, one major part essential for each species member to exist, and the other minor one which can vary within limits and determine differential survival in changing external conditions. In agreement, stasis is readily observed in the fossil record as oscillations of morphological characters around the stable mean over time. If stasis is to be explainable by system information, the latter must be stable--unchangeable during the specie's lifetime; it is born and dies with the species resulting in stasis. The first, the last, and all members of species in between, have it and exhibit the "same" species phenotype on that basis. Simply put, this postulate means that microevolution does not extrapolate into macroevolution.

Our theory holds that evolution makes species-specific system information packages (Crkvenjakov and Drmanac 2007, Heng 2009, 2019, Crkvenjakov and Heng 2021). New species rarely arise. The rarity attribute applies to a human life scale, making these positive outcome events unobservable in nature. Still, they are nevertheless sufficiently frequent (on a scale of millions of years) to account for the advent of all long-living species having life spans longer than certain thresholds, say one to a few million years, that ever appeared on the planet (Serpkoski 1998). These are the overwhelming majority of all fossil species,species of microbes and a certain number of currently existing species whose longevity above the threshold could be ascertained. Molecular clock studies of genomes of prokaryotes and protists indicate even greater longevity of their species than species of multicellular eukaryotes.

Needless to say, evolution concerns itself only with long-living species. The younger extant groups are unconfirmed candidates for evolutionary relevant species status. By pushing the paleontological temporal resolution by two orders of magnitude to the new remarkable record, Crampton et al. (2020) recently confirmed that among the groups that could be detected in the extinct graptoloid clade over its almost entire 62-million-year lifespan, the ephemeral species with durations of 25,000 to 500,000 years minimally comprise about half of all species, while the rest are the long-living species over a million years old. Species lasting less than 25,000 years were not included in the above tally. Our own very young *H. sapiens* group is almost certain to pass muster by special pleading based on its ecological success and the complexity of its novelty in respect to related groups and species. Still, many other extant groups, perfectly good species by criteria other than duration, could be just ephemeral oscillations around the phenotypic mean within the already existing encompassing system information package or simply future evolutionary failures/ephemeral species with the short life expectancy.

Treatment of molecular information in biology, and especially its analogy to computer software in molecular biology, has long been criticized as naive and untenable by information scientists, physicists, and philosophers (Longo et al. 2012, Stuart 1985, Godffrey-Smith 2000). Further, no attempt at integration from critics from quantitative sciences appears to be mutually acceptable to molecular biologists. Rarer still is a viewpoint that there is a way to accommodate both sides of an information debate as substantially correct. This is the stance of late biophysicist Werner Loewenstein (1999, 2003). His view of information in biology is based on chemical thermodynamics, but information theory is introduced by his pointing out the equivalence of Boltzmann entropy and Shannon information as well as the importance of Shannon's channel capacity and coding theorems for biology. His guiding principle in discussing information in the molecular world is the requirement for its minimization in specific interactions. Loewenstein believes his principle of information economy is the potential Occam's razor for sorting hypotheses in theoretical biology; the process requiring the least information is the one that took place. For the emergence of life, he proposed that it happened by his conservative selection and not by Darwinian natural selection. In contrast to the latter, where the loser in the competition loses totally body and information both. In the former selection, the loser's information is taken up and lives on in the winner as the cheapest way for the winner to survive while the rest perish. His view of biology is that organisms stay the same and evolve due to

 doi:10.20944/preprints202210.0371.v1



information. The first aspect is covered by the concept of circular flow of information through organisms. In circular flow, inputs of information from outside are ultimately from photons; output is entropic dissipation to the outside, but informational macromolecules acting as Maxwell's demons sucking on photons accumulate prodigious amounts of information during the organism's lifetime so that reproduction can be achieved by closing the circle. Each bit of information helps the overall result to keep organisms far from equilibrium. The dogma of molecular biology of information flow from DNA to protein is only a segment of the circular flow, as the information flows back from proteins to make nucleic acids to close the circle.

Loewenstein answers the two basic questions about what evolution is and what evolves in the following way:

"If growth of information is essence of evolution, so is cybernation, and self-reproducing loop achieves the ultimate saving: total conservation. The number of cybernating loops increases at every evolutionary stage and loops form loops within loops within loops." (p95) and "The complex became twisty and labyrinthine, but all through evolution it maintained its circular shape, for it grew by piling up information circles, more than piling up information in each circle." (p109). His term for the cybernetic complex of information loops is circus[1]. What evolves is circuses, and they do it by growth of information. Hence, circus is the cybernetic equivalent of the species system information.

The three topics of interest to this paper have been quantitatively modeled by biocomputation scientists previously. In contrast to our information-based treatment, the issue of advantage of molecular sex over lineage accumulation of mutations can also be approached on the basis of algorithmic efficiency. In his book *Compositional Evolution*, Watson provided proof by mathematical modeling for the inability of natural selection to accomplish the creation of complex solutions in a single lineage that recombination or merger of parental genomes from different lineages can, assuming that genetic material is organized in interdependent, but semi- independent, modules (Watson, 2006). Also, in the context of a stability of "a given species' genomic complexity," in order to explain Drake's rule of constancy of mutation rate per genome over vast evolutionary distances, Schedrin and Parkhomchuk have mathematically modeled the possibility that the fixed channel capacity of species genome DNA severely limits the extent of allowed harmful mutations (2012).

Recently, Lois and his group (Johnston et al., 2022) have demonstrated in quantitative analyses evolutionary bias favoring simplicity and its symmetry variant and discouraging complexity in large phenotypic data sets, like tertiary structure of proteins, secondary structure of RNAs, and organization of GRNs. While properly identifying the source of this bias as random mutational change and thus excluding any role of natural selection in its generation, they seem to place its cause outside biology into the general mathematical nature of algorithms. However, we believe that the fundamentally parsimonious nature of evolution is expressed through our law of information conservation and that simplicity and absence of randomness, observable only in biology, is the consequence of this law.

[1]circus = augmentative of circle (lat.)

## Results and Discussion

*Biological Information: Symbolic, Sequence and Structural Information*

It is necessary to extend Lowenstein's treatment of relevant biological information as entirely structural to include symbolic information as well. This extension means entering uncharted territory of information theory outside its main confines. It concerns its relations with physics, philosophy, and in this case, biology, or how to interpret information rules concerning the transmission of unidimensional (linear) sequence of symbols through a communication channel. Three interpretations possibly suitable for the purpose of our discussion can be provisionally named physicalist, intermediate and semantic[2]. In the first, symbolic information does not exist, only structural information does, or the former

 doi:10.20944/preprints202210.0371.v1



has no further meaning than the entropic description of a physical object of linear sequence. All that is relevant is the measurement and interpretation of mutual information of a sequence (Adami 2016). With this approach, for instance, it makes equal sense to measure this parameter of both genomic nucleic acid and proteins, thus disregarding the special templating information relationship between them as biological molecules. In the intermediate approach, the conventional structural information is split into two parts. First, potential information, which through recording can lead to symbolic information, which is the only actual information, is always conveyed by a sequence of symbols. Second, the "proper" structural information of multidimensional objects is achieved by the irreversible semantic action of symbolic information. However, this symbolic information conforms to all Shannon rules without any reference necessary to its meaning (Battail 2014). The semantic approach treats symbolic information as emergent from the structural information of sequence. New rules, like code invariance and symmetry, protect and maintain meaning of the message conveyed by the sequence carrying it, but Shannon's quantification applies only to the structural information of physical embodiment of sequence accompanying symbolic information (Feistel and Ebeling 2016, 2011). The main argument favoring this interpretation is the fact that the unique message of the same meaning can be transmitted with various sequences having different Shannon entropies due to this many-to-one relationship, depending on language, coding, symbol choice, etc., or conversely of the two sequences with the same entropies, one can be pure nonsense, [2] The term "semantic" is used in a broader sense, then generally understood, to include natural messages which do not fully satisfy the criteria of meaningfulness, intentionality, aboutness and veridicality (Floridi 2010). while the other caries the specific message. While Shannon channel capacity and coding theorems must be obeyed qualitatively in terms of efficiency, the actual quantification seems not to be directly applicable to the fundamental biological questions pursued here. Relevance to biology of information as a "meaning" in respect to other forms of information has been suggested (e.g., Harms 2006, Crkvenjakov and Drmanac 2007, Koonin 2016).

While the stability of genetic code seems on the surface to go against the grain of the main argument of the semantic approach, the code itself provides the possibility of different sequences conveying the same message through its redundancy. Further, modern genomics has supplied enough examples of loops or their parts of circus/system information that have changed their genetic DNA sequence substrate during evolution while keeping the function/message intact to make it appealing to provide a framework for understanding this otherwise incomprehensible phenomenon.

Crucially for this work, proponents of the intermediate and the semantic approach both agree that symbolic information emerged only with life and exists now only within it and its derivations, and that copying of symbolic information does not increase its quantity or quality. However, by confining themselves to sequence-associated information only by agreeing that structural information is merely a potential one, they miss the chance to fully integrate information into biology. Following our description of information traffic in living systems consisting of an interaction network and instruction parts (Crkvenjakov and Drmanac 2007), here we distinguish three kinds of information: one associated with the interaction network exclusively and the two comprising the instruction component. We reserve the term Structural Information for information making possible mainly specific molecular bondings of informational macromolecules with others of their kind or with smaller molecules (but cf. lipid bilayer, etc.). On the other hand, the biological instruction part consists of sequence information and symbolic information. The first concerns the entropic aspects of sequence, is a specific (digital) kind of structural information, is quantifiable in bits and is subject to Shannon theorems, and the second is concerned with the meaning of the message when the meaning of the instruction emerges from the nonsense sequence containing the same or similar sequence information. Symbolic information is only indirectly quantifiable via its essential symbiotic link to sequence

 doi:10.20944/preprints202210.0371.v1



information, making biological Information Laws expressible only in probabilistic trends and not in exceptionless instances.

*Law of Conservation of System Information*

> *Living systems evolve in speciation by following the path that minimizes the cost of their required symbolic information growth, while the (nonreductive) offspring species simultaneously maximally conserve the symbolic information of their parent species.*

Probability laws of physics and mathematics are based on minimization of variables under consideration. The difference of treatments of information in chemical thermodynamics and information theory allows resolution of the fundamental duality of living systems to have the ability to both stay the same and diversify for a time with the conservation of information law. This law says that if one considers two states of the living system differing in internal information content, the pathway connecting the lower to higher information state will be inevitably chosen among alternative ones that require the minimal increase in internal quantity of information.

Loewenstein's principle of information economy is: "(1) biomolecular systems tend to minimize information expenditure; (2) they evolve along paths maximizing information efficiency (Loewenstein 1999, p 91)." From the context one can discern that (1) the principle applies to single systems and (2) to their competitive interaction in the sense of who wins if systems vie for the same time-dependent goal. For Loewenstein, the information transactions in prebiotic and living systems are governed by the thermodynamic equation [his eq. 3, RC]: $\Delta$ Im=$\Delta$ Sm + $\Delta$Se, where I=information, S=entropy, m=molecule, and e=environment, which states that any increase in information of a molecule has to be more or less instantly balanced by an increase of either the entropy of a molecule, or entropy of environment or both. It is a difficult condition that biomolecular systems manage not to violate by minimization of the required information ($\Delta$ Im).

To cover similar ground, we independently proposed the lock-in law: "[A] particular program solution for any or all aspects of species phenotype, once directly or indirectly found to be necessary for the survival of an evolutionary lineage (clade), has to be retained, modified or not, in that lineage in all its divergent branches. In short, it [the lock-in law] mandates that evolution proceeds along the path that maximizes the conservation of existing instruction" (Crkvenjakov and Drmanac 2007, p55). As our formulation is addressed to pattern the evolution of life once it has emerged, it is less general and more predictive. We held that instruction is encoded information in DNA (symbolic information here) and that (species) system information consists of instruction and the cellular interaction network dealing with information transfer. The only portion of information excluded then, but not here, was the one dealing with the "business of living" functions supplying metabolic resources and cellular structural support functions which are remarkably constant over many lineages in evolutionary time.

Loewenstein's formulation might be considered to answer the question of what one can say about the increase in the content of information of the offspring species in respect to its parent in evolution and ours what proportion of parental information is retained in the progeny. Assuming a given fixed amount of symbolic system information must be attained consisting of the sum of parental and new information for any direct progeny species to exist, the additional amount of information required can be minimized when the retention of parental information is maximized. The two formulations are therefore two (verbal) sides of the same coin and are united here in the single Law.

The consequence of law is the conservation of speciation achieved gains of symbolic information in stasis phase of species' life and equally in a species lineage as a whole. As

 doi:10.20944/preprints202210.0371.v1



locked-in by system information, the package emerged at speciation, a function dependent on this symbolic information cannot be lost on pain of inviability except in two cases of functional replacement in further speciations in a lineage. In the first case, there is a gain of new symbolic information which codes for the same function but by a different underlying mechanism allowing the equal loss of old information. In the second case of reductive evolution, a species loss of a given function occurs when it is already compensated by the function of a mutualistic partner species.

Note that the conservation law is not neutral in respect to the question "Is evolution gradual?" as it requires saltational growth of system information in each successful speciation resulting in a novelty, regardless of this growth being of minimal information size possible to achieve this novelty. The statistical possibility distribution ensures that on rare occasions this size is large enough to allow recipient species to become founders of higher taxons as required to account for the biological diversity pattern.

*Information in Chemistry:*

Nucleic Acids have Unique Information Transfer Capacities

In the world of water-soluble organic molecules of certain complexity to which life belongs, information is contained in molecular structures. The only natural way information can increase or decrease in the internal molecules is through its transfer upon bonding with the other molecules in the system (apart from photon and other radiation absorption). The less random and more specific these reversible bonding interactions are, the greater the information transfer is. In other words, the transfer is proportional to the number of weak chemical bonds the molecules can participate in during binding, and that is the same as the degree of spatial complementarity of their atomic contact surfaces. To exist longer than an instant, a system must be capable of a set of interconnected bondings/interactions of participating molecules among themselves, or resources from outside, resulting in a flow of information. The larger the number of bindings and their individual specificity is, the larger the amount of information in the system. As we consider systems that can last as they are for an appreciable time by pumping in information from outside to maintain the achieved level of interconnected bonding and also increase information content by advent of new bindings to achieve more complex chemical composition, the law imposes an additional selection condition to achieve bonding at the minimal information cost of the least bits possible. This condition might have sufficed to uniquely select, as constituent molecules of the first system with rudiments of stability far from equilibrium and capacity to grow internal information, the classes of molecules we see in cells today, like nucleic acids, proteins, and lipids. While science cannot exclude a possibility that life started with a stage consisting of systems of some other molecules (Joyce 2012), the law provides a stringent chemical criterion of requiring these molecules be similarly proficient as current life molecules in bonding each other specifically with as little information.

Life as we know it is based on nucleic acids. The chemical capacities of nucleic acids are exceptional, or even unique, compared to other natural molecules. In order of decreasing familiarity, these chemical capacities are: (i) unlimited capacity for storing symbolically encoded information in sequence of its monomers, (ii) unique property, known as "molecular magnetism", or ability to promote strong bonding within itself and of other molecules either of its kind or other kinds specifically through formation of complementary base pairs, or generally, by offering enormous sufficiently variant contact surface, and (iii) propensity to lengthen significantly the sequence by either end-to-end fusing or recombining into one the fragments of two nucleic acid molecules of different sequence thus potentially increasing the information content of a resulting molecule. Aside from those RNAs which need spatially proper configuration for function, nucleic acids, which are in the first approximation pseudo- linear in cells or viruses, do not carry much structural information. Conversely, among life- generated molecules, nucleic acids, although

 doi:10.20944/preprints202210.0371.v1



structurally simple, are uniquely rich in symbolic and sequence information contained in their sequences. The enumerated capacities have fundamental importance for molecular information transfer and content. It is our assertion that the advent of nucleic acids during the course of chemical evolution provided the necessary and sufficient capacities for the emergence of life in appropriate conditions otherwise sterile without them.

*Life is Co-emergent with Species, Molecular Sex, and Code*

Speciation Law of System Information

> *Life emerged by molecular sex of abiotic systems producing the first alive system information/circus containing organism/species. Henceforth, in descent by modification from this first species ancestor, species arise by horizontal molecular sex between organisms from maximally possible divergent vertical inheritance lineages.*

It is usually said that first organism consisted of an RNA molecule or molecules that could have replicated themselves using the required monomers and energy as externally available resources. The unspoken assumption is that complete replication suffices to form two organisms where previously one existed. However, the capability to divide physically is separate from replication and could have emerged in the first organism in parallel with the ability to close the loop of reactions required to replicate its RNA contents, or this capability could have been already present in simpler abiotic parental systems. If the latter is more likely from the information gained as well as from experimentally demonstrated standpoints, then the fusion of two systems as reaction reverse to division must have been possible for systems parental to the first organism as well. This fusion is made likely, if we suppose that a replicators-in-a-compartment-capable-of-division system is not alive without acquiring template transcription function as well (Szostak et al. 2001). Fusion of two vesicles, one containing the system and the other either containing another such system or only inert RNA molecules, is the simplest way to envision enrichment of RNA content of pre-life systems that could have led to life. The nucleic acids' capacities ii) and iii) from above are responsible for the ability of nucleic acid systems-- both prebiotic and living --to engage in recombination and ligation among internal and exogenous nucleic acids in the process we call molecular sex.

Thus, RNAs originating in other systems could have participated in information transfer reactions of the system in which they found themselves due to the unique capacities of nucleic acids for "promiscuous" polymer chemistry.

The information the receiving system gains by engaging in molecular sex is free, since the information cost of recombination and ligation with exogenous RNA as well as initial information gain for the system is zero bits. As mentioned, Leowenstein championed this process in his conservative evolution proposal: "However, raiding, the takeover of information from another system is permissible under equation 3. Living systems resort to this practice.... The information in either system had been correctly paid for before the merger" (Loewenstein 1999, p89). Unsurprisingly in extant cells, products of such recombination and ligation, at least in some cases-- if not generally, are replicated and passed on to next generation and down the lineage, as reproduction has forever erased the distinction (within the system) among host and foreign origin of new nucleic acids sequences gained in the molecular sex. One very sharp caveat is needed: the assertion above is valid only for the previously exogenous nucleic acids which have become part of the genome in the cell. Provided that the new sequence in the template or its protein product can bond any molecule and thus engage in information transfer in the cellular environment of acquiring cell, the host offspring would have gained symbolic information as well as appreciable bits of sequence information from its parent at no cost, i.e., the absolute minimum possible as required by the law. The genome location ensures the information gain will be stable; in other words, it will be passed on to a new generation, the daughter cells and

 doi:10.20944/preprints202210.0371.v1



their progeny. Consider a bacterial cell acquiring a new antibiotic resistance gene by horizontal transfer in an example from the extant life. If a foreign gene is spliced into a chromosome and has regulatory signals recognizable by host cell machinery, the gene is used to produce a new protein required for the cell and its progeny to survive in the antibiotic-containing environment. This molecular sex trick for bypassing entropic harshness in exacting stern payment for every bit of information is uniquely open to molecular systems having nucleic acids functioning as the genomic template. On questions of the origin of molecular sex, one should regard two stages in respect to information gain as in Loewenstein's model (1999). The first in which there is no symbolic information gain either because the system has no template or because once part of a template, the new sequence elicits no binding reaction in the system and thus has no function. Expanding the bit content of the system genomic template nucleic acids with functionally inert sequences (although without initial gain of symbolic information), can be beneficial, as it represents an additional target for a mutation that could make it functional after some generations. This kind of molecular sex could have originated within the first systems that acquired replication function as it does not require template function. The second stage with symbolic information gain could have originated with the offspring of the first system, which had the template function, or at any point thereafter.

However, it is tempting to propose that life emerged as a consequence of the first act of molecular sex with symbolic information gained in the progeny. This hypothesis would require that template function was missing in nearly alive prebiotic systems and that life emerged when two of them fused, and one of the products of recombination and ligation of their RNAs took the role of transcription template, and its complement chain the role of phenotype, while the system retained replication and compartment function as well.

Speciation law answers the simple question, “What came first in evolution- gene or genome?” by asserting they appeared simultaneously in the first species progenitor of all life. Contemporary science never posed this question due to *a priori* assumption that “co-operating” genes came first—an assertion which cannot be supported nowadays by any valid argument.

The universality of horizontal transfer today and its involvement in the emergence of the most significant novelties seen in evolution leads to an argument of universality of molecular sex in speciation and the question of when this process made its appearance. The simultaneous emergence of template transcription and selectable phenotype of the ribozyme in the first cell by fusion of two abiotic systems possessing RNA replicase function is a proposal that can be defended since it features simultaneous closing of the feedback-based circular loop of information flow and the start of the first species lineage. Although epistemically one can call the template for this ribozyme the first gene, parsing genomes into genes does not generally carve nature at the joints, as the latter are undefinable as a unit of function in most organisms.

The argument for validity, if any, of this solution, applied universally to origins of species onward during the history of life, lies in its ability to answer several questions at the same time:

1. How living systems maintain and accumulate information to exist far from chemical equilibrium for billions of years?
2. How and when information flow in living systems split into two streams with separate roles in accumulation and maintenance of information, that is when genotype and phenotype emerged?
3. Is it and why is the horizontal transfer of inheritance by molecular sex a universal property of life?
4. What is life and to be alive, that is when did the first system information/circus emerge? Our answers are provided below.

*Species Novelty Universally Originates by Molecular Sex*

 doi:10.20944/preprints202210.0371.v1



We follow here the theoretic scenario that life emerged when an abiotic protocell system acquired a coded ribozyme, for instance, for the synthesis of fatty acids (Szostak et al. 2001) as this scenario is suitable to illustrate the informational transactions in accordance with the laws we propose. Specifically, in our variant of this scenario involving the fusion of two protocells, the combination by recombination and ligation of two previously inactive sequences coming from different parental systems in a fusion event could have produced the active sequence of a ribozyme via template transcription because that is informationally a more economical way of appearing than by the accumulation of information within any single system.

Cladistics holds that a new species distinguishes itself by possession of a specific novelty absent in its parents (Eldredge 1989). More generally, molecular sex with horizontal transfer among unlikes provides a greater chance for finding functionally productive combinations of mutations than is possible by selective or non-selective accumulation ofmutations by vertical inheritance in a single lineage. Not only are accumulated mutations from the two independent sources combined, but also the very act of recombination and ligation can create additional mutations impossible to survive previously. Very likely, genome chaos-like drastic reorganization of the genome is required for a successful viable outcome of sex of sufficiently different parental genomes. From the information standpoint, the difficult price of gaining symbolic information by accumulating mutations is entirely avoided as in the supposed example above, where the active product is made of inert sequences acquired earlier by systems by the first stage molecular sex. Or, for instance, halved in the case when two previously functionally active sequences from the fusing systems contribute to the product with a novel function. The latter case could have represented the situation where a novel function required so many underlying bits that it could never happen by accumulation alone, either absolutely or because it would take such a long time that the same sexually produced functionality would appear first and thus preclude the survival of the latecomer. In any case, The conservation law requires that the found solution is as straightforward and simple as possible.

Even more generally, species novelty once acquired by the organism progenitor of a new candidate species is tested in the real world in a multitude of copies over generations of individual group members. Due to the rarity of new speciations, only the new symbolic information, captured by the novelty, that enables the group to survive above the longevity threshold of millions of years or more can be passed as copy of the original one on to the new species as a part of parental contribution to the total symbolic information store of the offspring species. In turn the next offspring species in a lineage, must acquire its own novelty which together with the parental symbolic information contribution enables it to pass the above test.

This is mandated by the conservation law requiring the maximal conservation of parental symbolic information.

*On Non-Randomness of Evolution*

Progress in genomics and structural biology in the last quarter of a century shows that the genomic sequences and 3D protein structures evolve neither randomly as creationists want, nor gradually as modern synthesis-based accounts hold, but in accordance with the lawful maximal conservation of existing information. New protein coding sequences arise mostly non- gradually by duplication or fusion followed by sub- and neofunctionalization, mostly by changes of a few residues. These processes are informationally cheap in decision bits since old, parental information represents most of the new. As a consequence, on protein 3D structure levels the present variation in domains is archived mostly by increasing in complexity of combinations of primordial structural fragments with evolutionary time (Bromberg et al., 2022, Wang et al., 2011). Currently, about half a million domains are known (from a few billion available protein sequences), and they are all related to each other by descent by modification, greatly restricting their structural variation in respect to random expectation. This relatedness is reflected in the level

 doi:10.20944/preprints202210.0371.v1



of multi-domain proteins in nested hierarchical classification trees, covering the entire organismal variation. Most (97%) of the 180k known protein structures belong to only three classes, within which, out of approximately 1300 known folds, nine superfolds contain more than 30% of all structures, and within which, out of 6100 superfamilies, 100 most populated ones contain more than 50% (Bordin et al., 2021). Plateauing of the discovery rate of the new folds despite the increase of numbers of structures seems to indicate the limits of natural structural variation. On the other hand, the theoretically possible folds vastly outnumber real ones (Taylor, 2020). Obviously, since this drastically curtailed variation compared to the mathematical one sufficed for the myriad enzymatic, and other functions proteins carry on in cells across the biosphere, it was possible for evolution based on the information laws to produce slight variation of these basic structures instead of being forced to invent further ones.

The mosaic domain nature of 3D protein structures and their surprising degree of relatedness demonstrates that base change mutations are restricted in their influence on protein evolution, while duplication and fusions of domains dominate[3]. Organisms are equipped to produce these informationally cheap, but productive, mutations in frequencies sufficient to account for buildup of structure complexity from a small number of primordial peptides to the present universe of complex, protein-based, molecular machines (Bromberg et al. 2022, Wang et al. 2011). This is a non-gradual, but lawfully step-wise process consisting of a sequence of the minimal information jumps. This does not mean that bit size of individual jumps in different times and in different lineages are the same. Size distribution covers the largest, so-called "frozen accidents," which are so rare that they occurred only once in evolution in the founders of the largest higer taxons to the smallest and most frequent jumps that connect the closest neighbors in a family or a genus. Lawful conservation protects the rarest informational gains most in inverse proportion to their frequency in evolution. This is accomplished in two ways.

Due to their longevity, these gains are viability protected by being entrenched in growing number of functional dependencies within each species, and, second, are spread over a number of lineages, thus preventing their loss by lineage extinction.

The law limits the successful products of the search for novelty in the vast sequence and structure combinatorial space only to those combinations that probe the regions in extreme vicinity of sequences and structures already proven to work in the real world. Numerically vastly dominating arrays of other possibilities do not need to be sampled for successful outcomes this search. Due to this, despite randomly directed mutations, mathematical randomness does not exist in evolution.

[3] Role of accumulation of neutral single base mutations in facilitating the positive functional outcome of the last one in a series via so called neutral networks (Schuster et al. 1994, Wagner, A. 2007) contributes to the lack of mathematical randomness as well but is not further discussed here.

*Molecular Sex and Other Evolutionary Processes in Speciation*

The proposed informational advantage of molecular sex in creating novelty in speciation should not be understood as an argument that vertical accumulation of mutations in lineages does not happen or can be dispensed with in the explanation of evolution. However, in our view, those processes consisting of natural selection and natural invention (accumulation of non-selectable mutation combinations in ephemeral, within-species lineages) operate below species level to make possible speciation by molecular sex (Crkvenjakov and Drmanac 2007). The natural invention process accounts for the required rarity of successful hybrids of unlike parents. If only natural selection accumulated mutation combinations, many more individuals from each parental group would carry these mutations, and molecular sex would be far more efficient in the sense of a number of hybrid organisms carrying the same combination.

However, due to the reduction in the number of different mutation combinations tested, relying on natural selection only would have disastrous consequences for the

 doi:10.20944/preprints202210.0371.v1



achievable levels of novelty. Natural invention mutation combinations being accumulated in genealogies within species in an entirely stochastic manner and consisting of nearly neutral mutations requiring activation in an appropriate hybrid context need numerous hybrid trails to find the super- combination of parents needed for novelty as seen by the neglected phenomenon of ample zygotic inviability present in animals including humans and in plants. This extremely low-odds lottery is imposed by the difficulty of gaining internal symbolic information for a new functionality. Parental mutation combinations accumulated by natural selection are required as well since the new species started by a successful hybrid needs to be preadapted to survive in external conditions that are closely similar to that of its parent species. However, natural selection is unable of eroding or eliminating functional modules of species-specific system information during lifetimes of non-degenerative species due to the inviability imposed by the conservation law.

Lineages on both microevolutionary, below species level, and macroevolutionary levels as species clades can exist because of vertical inheritance. Neither horizontal inheritance by molecular sex nor accumulation of mutation processes could take place without vertical inheritance. In other words, evolution is not possible without lineages. This reality argues against proposals for an evolution stage preceding the current one in which lineages were not needed (Woese 2002). Conversely, diversification into species lineages is both necessary for and an unavoidable consequence of the evolution of life from the time of origin. This argument cannot be defended without proper understanding provided in the section on system information bellow of the role of species both as a container of lineages of organisms and as a discrete unit in macroevolutionary lineages.

*Emergence of Code Simultaneously with Life*

The first ribozyme in Szostak et al. (2001) hypothetical emergence-of-life model provided the catalytic activity as a selectable phenotype to the system, just as the ribosomal RNA transcription provides the peptide bond synthesis ability to cells today. The ribozyme making qualifies as symbolic information transfer, similarly to transcription-translation today. In other words, the communication channel emerged that connected genotype and phenotype, and with it, the first code--ribozyme code. By proposing here for the first time that the ribozyme code preceded genetic code, we part company with most biologists and philosophers of biology, who would consider any template synthesis of DNA as transfer of genetic information (Stegmann 2005), or in our parlance of symbolic information, and that a code is possible only as a connection of two different chemistries--nucleic acid and protein ones.[4] The first symbolic code consists of the base pairing rules. The same rules are used for making unique spatial conformation of single-stranded RNA with the binding ribozyme function as for making a copy of the same molecule in replication. In the first case, the linear sequence encodes by "arbitrary" base pair rules a molecule active in information trafficking through its richness in structural information contained in its unique spatial conformation. Base pair rules act as a ribozyme code. There is no code in base pair rules in replication since it is strictly copying a process with no appearance of new meaning, interaction or complex structure in the system; thus, there is no increase in symbolic information (see more below) in replication alone today, and its content was 0 in the prebiotic systems before the emergence of the first ribozyme as symbolic information is contained only in the precise template for spatially active and selectable molecules. In basic chemical terms, replication of RNA is the formation of a structurally simple secondary structure of a double helix independent of a precise sequence of bases; the algorithmic rule specifying helix is simple - its structure is poor in information. On the other hand, synthesis of a single-stranded chain of the specific ribozyme spontaneously leads to it assuming a unique spatial configuration with its active center capable of binding small organic molecules and changing their covalent bonds, thus repeatedly pumping structural information into organism.

 doi:10.20944/preprints202210.0371.v1



These events are absolutely dependent on sequence; the algorithmic rule specifying the particular ribozyme is complex - the template structure is rich in symbolic information--since the code is needed to transform the random polymer chain structure to the exact structure with the precise spatial disposition of all chain atoms. In extant life, DNA synthesis is never transacting in symbolic information, while synthesis of spatially active RNA (and proteins) almost always is. One can say that life emerged as a durable materialization of an abstract chemical possibility- polymer tertiary structure.

[4] However, there was a recent mention of coding RNAs in the RNA world (Inkpen and Doolittle 2016), or "meaningful" transcription (Koonin 2016).

Information theory might furnish a deep mathematical reason why ribozyme code must have existed in the RNA world and why both the ribozyme and genetic code exist today; in the words of Battail (2014):

> *Let us more closely consider the case where a message is associated with a 'recipe,' understood as the sequence of instructions needed for constructing some object. Then a bit or a block of successive bits of the corresponding information message instructs a step of the construction process, hence its semantic content entirely depends on its location in the message. The steps of the construction process must be executed successively, so their order in the message must match their order in time. The time axis is a unidimensional space, and so should be the medium which bears the message. A recipe, thus, cannot be borne by a medium having more than a single dimension.*

Battail then goes on to assert that structural information cannot be transformed back into symbolic information since multidimensional objects cannot serve as a recipe, providing information theory rationale for the central dogma of molecular biology (2014, Yockey 1992).

*A System Information Package/Circus is the Essential Attribute of Species*

Definition of Circus

The stable informational entity that defines species in a new way can be described alternatively in molecular genetics terms as system information or in cybernetic and informational terms as circus. Its properties of long-lasting stability and the specific way of the stability interruption and change in speciation mandated by the conservation of information law promulgated here, explain how teleology (also known as function, purpose, apparent design) in biology is born within the first species and persists throughout evolution.

Following Loewenstein, we posit that information in living beings flows through a circus-a complex of interconnected positive and negative feedback loops, i.e., cybernetically. The flow is ultimately circular as it starts (arbitrarily chosen) at the birth of an organism and ends there with the birth of offspring as a result of self-reproduction. Once the circular flow occurs and thus completes a circus, from one molecular system, in the simplest case, two are made, each containing the same circus. Cut any one of the constituent loops in the complex of loops by preventing the previously possible binding interaction of constituent molecules, and circular information flow cannot be completed; the pathway to self-reproduction is blocked, and thermodynamic information decay (death) is inevitable.

As loops in the evolutionary earliest organisms always started by message transfer from template nucleic acids to phenotype consisting of spatially complex polymers (RNA or proteins), the growth of symbolic information in evolution is the rate-determining factor of growth of overall information, as sequence and structural information cannot grow without it. Without emergent symbolic information as part of the circus loop fixing the meaning of the novel phenotypic function, the function cannot be propagated through

 doi:10.20944/preprints202210.0371.v1



generations. Thus, any increase of information content by changes in sequence or structural information without changes in symbolic information, as for instance, caused by neutral mutations, is transient and of no evolutionary consequence.

In terms of a quantifiable portion of information consisting of structural and sequence information needed to sustain flow through circus, the amount varies with time as a system moves through all stages necessary to accomplish self-reproduction. As all individual loops of a circus are not simultaneously present at any moment in the life cycle, but appear as a result of information flow through the preceding loop and disappear when flow through them ceases, the content of the three types of organism information --molecular system--rises and falls with time in a reproducible manner within members of a given species. The circular information flow is directional. Due to it, the concept of circus is compatible with the dogma of molecular biology, which is directional as well. The obligatory flow of information from DNA to proteins is just a segment of the circle, which is completed with the flow of information from proteins to DNA in a different segment. Loewenstein (1999) did not see the paradox that appears in such presentation as dogma explicitly forbids flow of information from proteins to DNA. The apparent paradox stems from differences of information concepts of chemical thermodynamics and information theory. For the former, any synthetic work stores information in resulting molecules; thus, two copies of genome DNA have exactly twice the amount of structural information as the original before replication. For information theory, the crucial question is if the two have the exact same sequence. If they do, they are copies, and copying does not result in any new symbolic information; the information contents of the original on one hand, and any of copies, on the other hand, are the same. This later concept of information is narrower. It deals with symbolic or sequence information only. For this reason, some information theorists claim that information does not exist outside the living world, and from it, derived human artifacts, and that physics has nothing to do with a realm of information (Battail 2014, Feistel and Ebeling 2011, 2016). The concept of information as molecular structure is broader as it includes information of both sequence as physical object (its bits) and the overall structure of the molecule. To resolve the paradox, one must note that the dogma concerns only symbolic information and sequence information and forbids them to flow back from proteins to DNA.

*Stasis: System Information/Circus Is Stable Because It Obeys Shannon Channel Capacity Theorem*

The overall positive feedback nature of circus means that from the first cell/organism onward, chemical activity of living molecular systems has been committed only to the goal of reproduction, specifically, to the mode that attempts exponential growth of system numbers. This overarching goal imposed by the systems' chemical potential is tempered only by the necessity to adhere to physical laws and the braking effects of negative feedback loops within the circus responding to environmental conditions (Loewenstein 1999). Ideally, this goal is best achieved if the same circus which imparts success in survival is transmitted without change to an expanding number of progeny through generations in a lineage of organisms. According to the information theory postulate that copying does not increase information, organisms in a genealogy formed through reproduction have the same symbolic information. The overall information integral over life cycle remains unchanged, as non-symbolic information is constrained by the obligatory symbolic segment of the flow through circus. The group of systems each carrying a copy of the same circus are the members of the same species. The capacity to evolve, inherent in living matter through changes in circus, imposes the need to compromise with the absolute stability of circuses. The compromise imposed is an equilibrium of stability of circus for a vast but finite number of generations, punctuated by its change in the organism with a potential to be an incipient stage of a new lineage. This potential is achieved by adding a new feedback loop or loops to the majority of the ones preexisting in the parental circus. This phenomenological argument for stability is bolstered by the information theory theorem of the

 doi:10.20944/preprints202210.0371.v1



communication channel properties regarding its capacity for information flow and its ability to tolerate noise. If capacity theorem applies to the part of circus concerning symbolic genetic information, and its associated sequence genetic information, so that the channel exists there, then the mandated fixed channel capacity limits the error in the transmission of sequence information to phenotype to the level which does not stop the information flow. If the quantity of sequence information needed to flow through the channel connecting genotype and phenotype, not to corrupt symbolic information is close to its maximum capacity in a viable circus, the mutations as errors cannot be allowed to greatly change the phenotype. Thus, individual differences in phenotype brought up by mutations do not bring survival in question if they are small enough not to disturb the underlying sameness. The specific nature of each circus fixes its genotype to phenotype channels, but unavoidable mutations have the least chance to reduce numbers of viable progeny if copying of the circus by reproduction is as faithful to the original as possible. Note that copying of circus is copying of potential for essential and necessary molecular interactions and not copying the exact sequences of monomers in informational macromolecules participating in them. The many-to- one relationship between the primary sequence and the final spatial structure of these molecules, as well as the redundancy of genetic and other biological codes, makes copying of species programs/circuses robust to a considerable part of errors in the chemical synthesis.

The stability of the communication channel is the result of the stability of genome DNA organization in the species during its million-years-plus long existence. This organization born in the speciation phase consists of many layers starting with the position of protein-coding sequences and extending with locations of regulatory sequences and up to sequences ensuring the entire chromosomal structure. The tip of the iceberg of this complex organization is the overall cytological karyotype picture. Species karyotype is not absolutely stable, but the allowed superficial changes are few and far between (Heng 2019).

As mentioned above, the communication channel of circus to restrict mutations in order not to exceed its capacity allows the accomplishment of the circular flow, i.e., conservation of circus through reproduction. The means of this restriction is the death of the individual with the mutations that overshoot the channel capacity. The inviability restriction is not absolute, as discussed above. If the combination of mutations changes the channel itself by creating a new loop of the circus, its lethality expected on the capacity theorem will not materialize. This process is the only allowed way for a birth of a new species by the stable gain of internal information, in the necessity of all three kinds. As chemical reactions cannot be made absolutely error-free, mutations must exist as departures from the ideal stability of genome nucleic acids through generations. Therefore, organisms containing the “same” circus cannot be identical but contain individual differences in phenotype based on those mutations that are allowed by the information capacity. Were self-reproduction ideally perfect to produce identical organisms, there would be no gain in information content when a group of species organisms is considered in respect to a single one. In reality, a population or a lineage (genealogy) of organisms within a species has higher information content than the organism itself. This greater information can be either ecologically or stochastically advantageous for building a successful combination of mutations for speciation. Populational gain of information is not used only for the evolutionary purpose of speciation, but also as a necessary buffer for species’ immediate environment tracking via natural selection of the survivable part of the inevitable change in external conditions within species’ lifetimes.

*Macroevolution Above Species Level*

Evolution as Growth of Shannon Codes

If our argument that a given species stasis is due to the copying stability of its circus is correct, circuses last as long as species--on average over a million years and maximally

 doi:10.20944/preprints202210.0371.v1



a couple of hundred million years. This timeframe is far shorter than the stability of life as it has lasted four billion years already and is still going strong. Can information theory account for the extended durability and the increase in complexity of life as well? We attribute life's longevity to adherence of living systems to the second Shannon theorem that proper encoding allows accurate transmission of the message regardless of its bit content and the extent of noise in the communication channel. The ability of life as a plurality of species to gain internal information and evolve since its single species origin is based on its commitment from the start to use a significant part of information flow in its circuses only to channel conveying encoded symbolic information efficiently. This commitment was absolutely necessary as, in a diffusion- driven, random, watery world of molecules characterized by incessant movement and collisions, the systems' capacity for information transfer solely by the specificity of interaction leading to change in chemical structure is strictly limited. For life's emergence and its increasing complexity with time--greater flows, even for evolution's open-endedness, a code (strictly in the information theory sense) had to be a part of the emergence, and later new codes had to be found and employed along the ones which have appeared earlier. This incorporation of code is mandated by the channel coding theorem, which grants to communication channels the unlimited potential for increasing transmitted, symbolic information with the application of appropriate coding. Some codes in order of appearance in evolution are: here proposed Ribozyme code (RNA-RNA, DNA-RNA); Genetic code (RNA/DNA- protein); and Cell-to- Cell (including Neural) communication codes in multicellular organisms (Loewenstein 1999). For the recent list of proposed biological codes, some of which might not fit the information theory definition of code, see Barbieri (2017). Mainly in animals, some circus loops are only indirectly connected to nucleic acids and may even consist only of cells, organs or even organisms constituting communication channels among themselves requiring further specialized codes. It seems possible to extend the stated information principles to these inventions, appearing very late in evolution and culminating in intelligence and culture. However, this extension is not attempted here.

*Life Persists By Sharing Gains Of Symbolic Information Among Evolving Lineages*

While the emergence of life is possibly lawful, its persistence is not guaranteed by any laws. For example, the total extinction of life is possible in case of the loss of planetary physical conditions which allow the existence of organic molecules. In its history, life has resisted with further diversification, either temporary, or less severe external conditions, which led to partial extinctions which resulted in the loss of some but not all of its lineages. In spite of those setbacks, a number of species probably expanded exponentially from the origin of life (see Benton 2016, regarding the last eighth of evolution's span for which such measurement is possible). If so, the open-endedness of evolution is possible. We suggest here that sharing of information gains by molecular sex among lineages is both inevitable and necessary and thus a universal feature of life. However, the odds of successful sharing are inversely proportional to both size of the informational gain and genetic distance among lineages, effectively prohibiting some transfers and limiting through sheer improbability the other important ones from happening more than once in the planetary time. For these latter cases, law mandating conservation of parental information ensured that the rare inventions, once acquired, were never lost. Life's evolutionary algorithm of accumulating information gains in independent lineages dominated by vertical inheritance and then testing them in a new context by mixing horizontally in molecular sex significantly increases the probability of successful outcomes over single lineage accumulation. The statistical edge thus afforded is sufficient to grow the Tree of Life since whatever is found to work successfully is retained in a new lineage and then used for leverage in subsequent pairwise trials with other species lineages.

There are two possibilities for sharing symbolic information by molecular sex. In the first, sequence information is shared among parental lineages in these events but not func-

 doi:10.20944/preprints202210.0371.v1



tionality as well. The latter emerges in the hybrid as new symbolic information from molecular sex by combination of nonfunctional parental components. Most of the smaller complexity novelties responsible for lower taxonomic categories are of this kind.

The second possibility is when both ready functionality and its information are shared either once, at the origin of lineage, or repeatedly from one lineage to start several. For instance, the mitochondrial oxidative phosphorylation was obtained by the symbiotic merger of the specific prokaryote with the host cell at the origin of the eukaryotic clade, but clade spreading of the novelty was achieved exclusively on the basis of vertical inheritance. In contrast, the original aerobic photosynthesis was repeatedly transferred (but not from the same donor) among lineages, as are the genes shared among species of bacteria, like in the already mentioned non-speciation example of antibiotic resistance.

Generally, information sharing is responsible for whatever capacity life has both for growth and speedy recovery after partial extinctions. One is entitled to doubt that life could have survived some of planetary upheavals or spread and expanded as it did if it were forced to diversify exclusively by vertical inheritance. Maybe life would still be lurking in the anaerobic mud waiting for its chance if mitochondria, chloroplasts and their cellular precursors had not made their appearance after oxygenation of the planet.

**Conclusion**

The law of conservation of Information holds that for two states of the living system differing in internal information content, the pathway requiring the minimal increase in the quantity of necessary information connecting the lower to the higher information state is always chosen among the alternative ones. If this gain happens in speciation, the daughter species is simultaneously kept similar, to the maximally possible extent, to its parent species in information.

Life, as we know it, is based on nucleic acids. Their chemical capacities are exceptional, or even unique, compared to other natural molecules. In a process we call molecular sex, these capacities are responsible for the ability of molecular systems, either prebiotic or cellular, to borrow and thus gain “for free” either, only meaningless sequence information, or simultaneously both symbolic and sequence information by the uptake and use of nucleic acids originated in other systems. The Speciation Law implies that molecular sex provides a pathway of the minimal information gain that preceded and was necessary for the emergence of life.

Henceforth, molecular sex capacity as a universal life property inevitably led to diversification of living creatures via hybridization of representatives of different lineages. In other words, the information shortcut possible by nucleic acid chemistry once found enabled not only the emergence, but the persistence of life as a descent by modification as well. Lawful universal reliance of life on molecular sex meant, among other things, that evolution could not eliminate viruses as obligatory mutual parts of its lineages from its very origin and throughout the history of life.

According to the conservation law the fundamental principle of (non-reductive) evolution is that the growth of information in species’ lineages proceeds in the most efficient possible way, in other words, by the minimal number of steps. Information as thermodynamic quantity is governed by the time arrow. The more bits of information needed to specify molecular interactions underlying an evolutionary novelty required for speciation, the longer it will take to accumulate them. In a race for time, the solution that requires the least bits will be found first and prevent the ones requiring more decisions (bits) to make the same result ever appear. Inversely, ruled by the law mandating conservation of symbolic information, is the need for the maximal possible retention of parental species information since the sum of the maximal parental and the required novelty information maximizes the information content of the offspring species. The latter pattern of the conservation of the functionalities based on the paternally achieved level of information is the omnipresent feature of evolution. For this reason, the most obvious inventions like genetic code, photosynthesis and mitochondrial respiration hold sway over the entire tree of life

or its large swaths. With this, the conservation law explains the elimination of mathematical randomness in, and the apparent teleology of, evolution.

Life can be characterized in alternative terms as comprising all-natural, reproducing, nonequilibrium systems with circular information flow. Instead of linear flows between Shannon senders and receivers, or reticulated flows in networks as in regulatory genetic networks, this flow is better visualized as Circus (Loewenstein1999). Circus is defined as an intricate complex of interconnected simple positive and negative feedback loops, each consisting of few molecular components, one of which is often DNA. The acceptance of circularity of information flow between the two domains of living systems, genotype and phenotype, precludes the prevailing use of a superficial metaphor of genome as the site of control of the cellular molecular system. However, for evolution and human intervention both, the route with the best chances of success in information input into living systems lies in the changing structure of genome DNA (sequence). As mentioned, DNA is the most frequent starting member of sub-component feedback loops of the circus as the source of symbolic information. Each natural circus is species-specific, or rather species, true in the informational sense, differ in their respective circuses. Such a concept of species represents the final move to transform biology to a mathematically based science from the science based on human perception and associated intuitions. It is uncertain what is the relation of such concepts to the concepts of species based on practical needs held by organismal biology. It is argued here that the latter concepts cannot support the universally valid theory of biology.

## References


Adami,C., 2016 What is information?, Phil. Trans. R. Soc. A 374, 20150230. http://dx.doi.org/10.1098/rsta.2015.0230

Bapteste E., O'Malley M.A., Beiko R.G., Ereshefsky M., Gogarten J.P., Franklin-Hall L, Lapointe F.J., Dupre J., Dagan T., Boucher Y., Martin W., 2009. Prokaryotic evolution and the tree of life are two different things. Biol. Direct., 4, 34 DOI 10.1186/1745-6150-4-34

Barbieri M., 2017. How Did the Eukaryotes Evolve?, Biol. Theory; 12: 13-26. DOI 10.1007/s13752-016-0253-3

Benton, M.J., 2016. Origins of Biodiversity, PLOS Biol., DOI:10.1371/journal.pbio.2000724

Battail, G., 2014. Information and Life, Springer

Berlinsky, D., 2019. The Social Set, Book Review, Inference, International Review of Science, Volume 4, Issue 4, https://inference-review.com/article/the-social-set

Bordin, N., Sillitoe, I., Lees, J. G., Orengo, C., 2021. Tracing Evolution Through Protein Structures: Nature Captured in a Few Thousand Folds, Front. Mol. Biosci., 10 May 2021, DOI 10.3389/fmolb.2021.668184

Bromberg, Y., Aptekmann, A. A., Mahlich, Y., Cook, L., Senn, S., Miller, M., Nanda, V., Ferreiro, D.U., Falkowski, P. G., 2022. Quantifying structural relationships of metal-binding sites suggests origins of biological electron transfer, Science Advances, DOI 10.1126/sciadv.abj3984

Crampton, J.S., Cooper, R.A., Foote, M., Sadler, P.M., 2020, Ephemeral species in the fossil record? Synchronous coupling of macroevolutionary dynamics in mid-Paleozoic zooplankton, Paleobiology 46, 123-135.

Crkvenjakov,R., Heng, H. H., 2022. Further illusions: On key evolutionary mechanisms that could never fit with Modern Synthesis, Progress in Biophysics and Molecular Biology. 169, 3- 11 https://doi.org/10.1016/j.pbiomolbio.2021.10.002

Crkvenjakov, R., Drmanac, R., 2007. Biological Evolution as Natural Programming, Academie Serbe des Sciences et des Arts Bulletin, Tome CXXXVI, Sciences Naturelles, publ. N° 45, 1- 108, available on request as pdf file from R.C. rcrk@comcast.net

Doolittle W.F., Zhaxybayeva O., 2009. On the origin of prokaryotic species. Genome Res., 19 5, 744-56. 10.1101/gr.086645.108.

Doolitle, W. F., 2019, Speciation without Species: A Final Word, Philosophy, Theory and Practice in Biology-quod.lib.mich.edu, Volume 11, No.014,http://dx.doi.org/10.3998/ptpbio.16039257.0011.014

Eldredge, N., 1989. Macroevolutionary Dynamics. McGraw-Hill, New York. Feistel, R., Ebeling, W., 2011. Physics of Self-Organization and Evolution, 1st ed.; Wiley: Weinheim, Germany

Feistel, R., Ebeling, W, 2016. Entropy and the Self-Organization of Information and Value. Entropy, 18, 193. DOI 10.3390/e18050193 www.mdpi.com/journal/entropy

Floridi, L., 2010, Information, A Very Short Introduction, Oxford University Press

Godfrey-Smith, P., 2000. On the Theoretical Role of "Genetic Coding", Philosophy of Science, 67: 26-44.

Goldschmidt, R.B., 1940. The Material Basis of Evolution. Yale University Press, New Haven.

Green, R.E.; Krause, J.; Briggs, A.W.; Maricic, T.; Stenzel, U.; Kircher, M.; et al. 2010. A Draft Sequence of the Neandertal Genome. Science. 328: 710–722.

Harms, W., 2006 What is Information? Three Concepts, Biol.Theory 1:230-242.

Heng, H.H., 2009. The genome-centric concept: resynthesis of evolutionary theory. Bioessays 31, 512-525. https://doi.org/10.1002/bies.200800182.

Heng, H.H., 2019. Genome Chaos: Rethinking Genetics, Evolution, and Molecular Medicine. Academic Press Elsevier, Cambridge, MA, USA, ISBN 978-012-8136-35-5.

Heng, H.H., Liu, G., Stevens, J.B., Bremer, S.W., Ye, K.J., Abdallah, B.Y., Horne, S.D., Ye, C.D., 2011. Decoding the genome beyond sequencing: the new phase of genomic research. Genomics 98 (4), 242-252.

Inkpen, S.A.; Doolittle, W.F., 2016.Molecular Phylogenetics and the Perennial Problem of Homology.J Mol Evol 83:184–192 . https://doi.org/10.1007/s00239-016-9766-4

Jablonski, D., 2017. Approaches to Macroevolution: 1. General Concepts and Origin of Variation. Evol Biol., 44, 427–450.

Johnston, J.G., Dingle, K., Greenbury, S.F., Camargo,C.Q., Doye, J.P.K., Ahnert, S.E. and Louis, A.A., 2022. Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution, Proceedings of the National Academy of Sciences 119(1), DOI 10.1073/pnas.2113883119

Joyce G.F., 2012. Bit by Bit: Darwinian Basis of Life PLoS Biology, 10: 1-6, e1001323

Koonin, E.V., 2016. The Meaning of Biological Information, Phil. Trans. R. Soc. A 374:20150065.DOI 10.1098/rsta.2015.0065

Loewenstein, W. R., 1999. Touchstone of Life: Molecular Information, Cell Communication and the Foundations of Life, Oxford University Press

Loewenstein W.R., 2003.Two Arrows from a Mighty Bow, in Gregersen, N.H., Ed., From Complexity to Life, Oxford University Press.

Longo,G., Miquel, P.-A., Sonnenschein, C., Soto, A. M., 2012. Is information a proper observable for biological organization?, Prog Biophys Mol Biol., 109(3): 108–114. DOI 10.1016/j.pbiomolbio.2012.06.004

Margulis, L., 2006. Bacteria Don't Have Species: Interview with Lynn Margulis, . Astrobiology Magazine, Oct. 6. http://www.astrobio.net/interview/bacteria-dont-have-species/ #sthash.M7wA47Xg.dpuf

Prosdocimi, F., Farias, S.T., 2021 Life and living beings under the perspective of organic macrocodes, Biosystems, 206, August 2021, 104445 https://doi.org/10.1016/j.biosystems.2021.104445

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L., et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468, 1053–1060.

Schuster, P., Fontana, W., Stadler, P.F., Hofacker, I.L., 1994. From sequences to shapes and back – a case-study in RNA secondary structures. Proc. R. Soc. Lond. B 255, 279–284.

Sepkoski, J., 1998. Rates of speciation in the fossil record. Phil. Trans. Roy. Soc. Lond. B353, 315-332.

Session A.M., Uno Y., Kwon T., Chapman J.A. et al. 2016, Genome Evolution in allotetraploid frog, Nature 538, 336–343 | doi:10.1038/nature19840

Shadrin A.A. and Parkhomchuk D.V., 2014. Drake's rule as a consequence of approaching channel capacity, Naturwissenschaften 101, 939–954

Shapiro, J.A., 2011. Evolution: A View from the 21st Century. FT Press Science, Upper Saddle River, NJ, USA.

Shapiro, J.A., 2021. Why the Third Way of Evolution is Necessary, Theoretical Biology Forum, 114, 2/2021, 13-26, DOI:10.19272/202111402002

Shubin, N., 2020. Some Assembly Required: Decoding Four Billion Years of Life, from Ancient Fossils to DNA, Pantheon a Division of Penguin Random House LLC, New York, ISBN 10:1101871334/ISBN 13:9781101871331

Stegmann, U.E., 2005. Genetic information as instructional content, Philosophy of Science 72 (3), 425-443.

Stuart, C.I.J.M., 2009. Physical models of biological information and adaptation, J. Theor. Biol., 113, 441-454.

Szostak J.W., Bartel D.P. And Luisi P.L., 2001. Syntesising Life, Nature, 409, 387-390

Taylor, W.R., 2020, Exploring Protein Fold Space, Biomolecules, 10, 193; DOI 10.3390/biom10020193

Wagner, A 2011. Genotype networks shed light on evolutionary constraints. Trends in Ecology and Evolution, 26(1), 577-584. DOI:https://doi.org/10.1016/j.tree.2011.07.001

Wallace, A. R., 1855. On the law which has regulated the introduction of new species. Annals and Magazine of Natural History, 2nd Series, 16, 184–196.

Wang, M., Jiang, J.J., Kim, K.M., Qu, G., Ji, H.F., Mittenthal, J.E., Zhang, H.Y., Caetano- Anollés, G., 2011. A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. Mol. Biol. Evol. 28, 567–582

Watson R., 2006 . Compositional evolution, MIT Press, Cambridge, Mass

Woese, C., 2002. On the evolution of cells, Proc. Nat. Acad. Sci. 99: 8742–8747

Yockey, H. P. , 1992. Information Theory and Molecular Biology. Cambridge: Cambridge University Press.