

# The Evolution of Life as a Categorical Information-Handling Process

**Antonio Carvajal-Rodríguez** <sup>1,\*</sup>

<sup>1</sup> Centro de Investigación Mariña (CIM), Departamento de Bioquímica, Genética e Inmunología. Universidade de Vigo, 36310 Vigo, Spain; acraaj@uvigo.es

\* Correspondence: acraaj@uvigo.es

## SIMPLE SUMMARY

In the world of Information Handlers (IH), entities absorb and process information from their environment, generating functional meaning as a result of their self-organized closure aimed at persistence and multiplication. Once used, that information is released in the form of observable patterns, which can be quantified through informational divergences. Within this world, variations arise in the ability to absorb and interpret information. Some of these variations increase the amount of shared meaning, fostering cooperation and the emergence of groups and hierarchical levels in information handling. This work briefly reviews two contrasting views of life and evolution: on one side, the reductionist or gene-centric perspective, which places the gene, as a replicator, at the core of life processes; and on the other, the relational or holistic perspective, which emphasizes interactions and systemic organization. Without taking a side in this debate, we show that by starting from the basic replicator model, but interpreting replicating entities as information handlers, one can study more complex scenarios associated with cooperation, symbiosis, and hierarchical organization. In such scenarios, the same replicator-type equations allow us to quantify the information generated by the handlers as they undergo changes linked to their processes of survival and multiplication.

## ABSTRACT

Living systems can be understood as organized entities that capture, transform, and reproduce information. Classical gene-centered models explain adaptation through frequency changes driven by differential fitness, yet they often neglect the higher-order organization and causal closure that characterize living systems. Here we review evolutionary frameworks, from the replicator equation to group selection and holobiont dynamics, and show that evolutionary change in population frequencies can be expressed as a Jeffreys divergence. In holobionts, this formalism allows partitioning the total information associated with selection acting on the genome and the microbiome, as well as the information involved in matching specific genotypes with particular microbiotas, yielding an informational signature of stable symbiotic coevolution. Building on this foundation, we propose a categorical model of Information Handlers (IH), entities capable of self-maintenance, mutation, and combination. Each lineage is represented as a discrete coalgebra with internal (M,R) closure, while the global system forms a supercategory of interacting lineages. Replication, reproduction, and transmission are modeled as morphisms whose informational changes are captured by a functor mapping biological transformations into transformations within an informational category, measurable through Jeffreys divergences. This framework connects gene-centered evolutionary dynamics with relational and organizational perspectives, offering a unified categorical view of living systems.

**Keywords:** life, evolution, replicator, information, category theory

## **1. INTRODUCCIÓN**

*What is life?* There are many definitions of life, approached from different perspectives: biochemical, metabolic, thermodynamic, and others (Margulis and Sagan 2000; McKay 2004; Gómez-Márquez 2021; Mariscal 2021; Skene 2024). At the most basic level, we can say that life is associated with self-organized molecular structures that are open to the exchange of energy and matter, capable of self-maintenance, renewal, and reproduction. These self-organized structures are what we call organisms. The process of life consists of the activities an organism performs to maintain its structure and function. In this work, we argue that life can be understood as a hierarchical system of information handling, where reductionist and holistic perspectives are not mutually exclusive but rather complementary for understanding its complexity.

When the process of life is occurring within a particular organism, we say that it is alive. For example, a tree is in a living state because it carries out processes such as photosynthesis and growth.

However, that tree is not isolated, it interacts with organisms of other species as well as with other trees. Fungi grow within and around tree roots, forming a symbiosis known as mycorrhiza.

Mycorrhizae are essential for the normal growth of trees because, among other things, they can absorb soil nutrients inaccessible to roots and transfer them to the tree. In return, the fungi receive the sugars they need for growth from the roots. As the fungal filaments extend through the forest soil, they may, at least temporarily, physically connect the roots of neighboring trees. The resulting system of interconnected roots is known as a common mycorrhizal network (CMN) (Lian et al. 2006; Beiler et al. 2012). Thus, there is interaction among equals (tree-tree) as well as between very different species (tree-fungus).

Undoubtedly, living beings depend on other living beings for their survival. For example, animals depend on plants for oxygen and food, and plants depend on animals and microbes for pollination and the decomposition of organic matter. In the biosphere, matter is continuously recycled. Decomposers such as bacteria and fungi break down dead organic material and recycle nutrients back into the ecosystem. This means that the matter necessary for life is reused and remains within

the system. The energy of the sun is captured and utilized by living beings through processes such as photosynthesis. Once within the biosphere, this energy is transferred and used in a complex network of interactions. This intricate web of interactions operates at different levels, and we can in fact think of it as a hierarchy of organisms within organisms, life within life. Viruses within a cell meet the criteria to be classified as living; the cell within an organism is alive; so are the bacteria within plant, animal, or fungal organisms; and the organism that contains those cells and bacteria is also alive. Finally, all organisms, all known life, exist, by definition, within what we call the biosphere.

The biosphere, therefore, is filled with life, or even, according to the Gaia hypothesis (Lovelock 1972; Lovelock and Margulis 1974), can itself be viewed as a living system that self-regulates and maintains the conditions necessary for life. But regardless of whether or not we consider the biosphere itself a living organism, it is undeniable that life is a process involving a complex network of interactions. These interactions are carried out by entities we call living, which interact with one another on the same level (as do cells within an organism or organisms within an ecosystem), but also across levels, both downward, with contained organisms, and upward, with the organisms that contain them. From this perspective, the organization of living systems appears not only as something to be explained (explanandum), but also as something that explains (explanans) (Mossio et al. 2016; Mossio 2024) .

On the other hand, life as we know it cannot be understood without evolution, the process that defines and shapes the diversity of life on Earth. Life is not static, living beings constantly interact with their environment and with other organisms, driving evolutionary change. Certain traits or properties of organisms change over time, and some of these changes persist and spread, whether by contingency or because they favor increased reproduction. In other words, the ensemble of entities changes over time, a process we call evolution of life on Earth. The capacity to evolve is also a key criterion distinguishing the living from the non-living. Viruses, for instance, lie at the boundary of the definition of life: they lack independent metabolism but do evolve. This raises a central question: *how does evolution occur?*

We can imagine a change, whether in the environment, the organism, or both. A change in the environment might be an unforeseen event, a natural catastrophe, or an accident, in which one class

of individuals dies while another survives or becomes isolated. Alternatively, after a change, an individual or an entire class of individuals may acquire an advantage expressed as greater reproductive success. If the change affects the organism rather than the environment, *how does that change persist over time?* It will do so only if the advantageous modification is passed from the individual to its offspring, and so on through successive generations. Although today we know of non-genetic forms of inheritance, and recognize their evolutionary importance, it is also true that even non-genetic mechanisms have genetic foundations that evolved through natural selection (Bonduriansky and Day 2018; Futuyma 2019). In other words, we know that the fundamental molecules enabling this process are nucleic acids, DNA or RNA, and the information they encode in what we call the genetic code: the set of rules by which cells translate that information into the proteins that build and maintain the organism.

We thus enter a second aspect of the living scenario. We began by emphasizing the holistic and relational component i.e. the multilayered web of interactions that constitutes the biosphere, but we now realize that throughout the long history of life, from the first cells, bacteria, or viruses some 3-4 billion years ago (Dodd et al. 2017; Westall et al. 2023), there runs a thread connecting that distant past to the present: nucleic acids. These are a universal feature of all known life and contain the encoded information that allows organisms to develop within their spatiotemporal existence. Whatever theory one adopts regarding the origin of the first molecules that prepared and enabled the emergence of life, all recognize the importance of nucleic acids as carriers and transmitters of genetic information. The transition to nucleic acids as a universal feature of modern life may have been a gradual process, in which the role of these molecules became increasingly central to the functions of life.

We therefore face two opposing focal points in the study of life and its evolution: one systemic or holistic, non-gene-centric, and the other reductionist. The holistic approach holds that the whole is more than the sum of its parts and therefore cannot be fully understood through reduction to its components. It studies life as an integrated whole of complex interactions forming an organized, dynamic, and deeply interconnected network across the biosphere. The reductionist approach, in

contrast, maintains that the entire system can be understood by analyzing its parts separately, especially the genetic information carried by every organism, which encodes the plan for somatic development and growth but also connects each present-day individual to the origin of life itself, forming a germline that stretches from the past into the future, where successive changes generate the dynamic and interacting diversity that constitutes the biosphere as we know and inhabit it.

In current literature, these two schools, holistic and reductionist, are often presented as sharply opposed (Sterelny et al. 1996; Griffiths and Gray 1997; Brenner 2010; Mazzocchi 2012; Nicholson 2019; Hancock et al. 2021; DiFrisco and Gawne 2025; Mazzocchi 2025; Noble and Noble 2025). Yet it is worth noting that even from a holistic viewpoint, the evolutionary perspective assumes that life began in simpler forms, which evolved into the great diversity we observe today.

One of the main hypotheses about the origin of life is the RNA world (Gilbert 1986), in which catalytic RNAs (ribozymes) self-replicated, albeit imperfectly, allowing optimization through natural selection. Although this hypothesis still faces several unresolved issues (Bernhardt 2012), it remains the most plausible one and serves well to illustrate the point we wish to emphasize: that at the beginning, living systems were simple enough to be approached from a reductionist perspective, and that evolutionary processes gradually generated more agency (interactions that constitute the organism itself and promote its survival and reproduction), more interaction, and more complexity (Newman et al. 2025; Shirt-Ediss et al. 2025; Moreno and Peretó 2026). This gradual complexification of life, from simple systems to interdependent networks, suggests that reductionist and holistic approaches are not mutually exclusive but complementary at different temporal and organizational scales. Thus, there is a bridge between reductionist and holistic perspectives: in the beginning, when organisms were simple enough, both approaches were not so far apart. Perhaps we could explain the whole by starting from the analysis of its parts and interactions, since the earliest changes likely arose bottom-up, from small individual variations. Although interactions with other organisms and with the environment, and the emergence of forms of agency such as secretion and movement, soon began to transform and direct the evolution of organisms, this process can still be viewed as driven by changes in hereditary material that enhanced those forms of agency.

The key idea here is that it is not unreasonable to begin the analysis of life's evolution from a reductionist standpoint, centered on a simple model of organisms whose dynamics allow for increasing complexity, both of the organisms themselves and of their interactions with the environment and with others, up to a point where the reductionist approach may no longer suffice, requiring instead a systemic or holistic framework. Nevertheless, the knowledge gained through the reductionist approach can be invaluable for developing a systemic model. In particular, genomic analysis can shed light on phenomena of major evolutionary significance, such as speciation (Payseur and Rieseberg 2016; Johnson 2022; Bock et al. 2023; Benítez-Benítez et al. 2025), as well as the ecological and evolutionary importance of symbiosis and symbiogenesis (Moya and Peretó 2011; Batstone 2022; Archibald 2024; Richards and Moran 2024; Wang et al. 2024), among others.

The goal of this work is to emphasize that one of the key elements of living systems is information handling. By handling, we mean the capacity to extract information from the environment, interpret it, and use it with functional meaning to enhance survival and reproduction. For example, quorum sensing is a mechanism by which bacteria assess population density and coordinate their behavior through the production, release, and detection of signaling molecules known as autoinducers. This process enables bacteria to act as an organized community, regulating gene expression in response to environmental and population-density changes (Miller and Bassler 2001; Ruan et al. 2026).

During the process of information handling, new information is itself generated, and this can be quantified. To show this, that organisms use information meaningfully for their multiplication, thereby generating quantifiable informational patterns, we begin with the simplest case: the genetic replicator. We then consider other units of selection, such as organisms carrying replicators, and show how informational patterns also arise when natural selection acts on populations of individuals, on mating pairs, on groups, and in cases of holobiont symbiosis.

In the final section, we attempt to move toward a more holistic approach, using category theory (Mac Lane 1971; Cheng 2022; Leinster 2025) to define a system that is organizationally closed upon itself, that is, a simple world of information handlers with self-generative and causally autonomous

organization, in which the informational fold (a measure of information changes produced within the system) quantifies the magnitude and functional significance of evolutionary change.

## 2 THE GENE REPLICATES, THE PHENOTYPE INTERACTS

### 2.1 Replicators

In evolutionary biology, a replicator is any entity capable of producing copies of itself, directly or indirectly, while preserving its structure (and the information that defines it) across generations, so that it persists over time and can undergo mutations or changes. This variation enables the action of natural selection (Dawkins 2016). According to Dawkins, the fundamental unit of evolution and the primary driver of life, is the replicator; although other authors have questioned the viability of the simple replicator as the origin of life (Yockey 2005). When we say that, under the action of natural selection, the world tends to become filled with more efficient or ingenious replicators, what actually happens is that those replicators, or more precisely, their vehicles (*sensu* Dawkins), that achieve a better fit with the environmental conditions in which they must replicate, do so more successfully. If that fit is maintained over time, those replicators will increase in frequency within the population.

But, *what does it mean to have a better fit with environmental conditions?* In the origins of life, for the first prokaryotes, it may have been as simple as greater efficiency in an oxidation reaction or greater stability in the face of environmental fluctuations. Later, in more complex systems, fitness could depend on abilities such as movement, predator avoidance, obtaining food, or attracting mates. The possibilities are multiple and depend on the level of organization considered, but from the replicator's perspective, all translate into an increased capacity to leave copies of itself.

In the context of population genetics, we would say that replicators with higher biological fitness have a greater expected reproductive success. The biological fitness of an entity can thus be defined as its general propensity to survive and reproduce in a given environment. This propensity reflects the expected outcome of its interaction with the environment through its phenotypic vehicle, the organism (Hansen 2018; Otsuka 2019).

## 2.2 Interactors

In defining biological fitness, we have deliberately used the term entity, since we refer both to replicators and to the units that express or carry them. The fitness of a given replicator depends on its capacity to persist and multiply within the environment in which its survival machine, the organism, operates. However, organisms carry multiple replicators (genes) that may not be transmitted together; therefore, the fitness of an individual replicator is not necessarily equivalent to that of the organism that carries it. Replicators constitute units of inheritance, but selection can operate at different hierarchical levels. Within this context arises the debate on the levels of selection (Lewontin 1970; Hull 1980; Okasha 2008), which distinguishes between (i) what interacts with the environment (the organism or the group), (ii) what replicates or reproduces (gene, organism, or holobiont), and (iii) what manifests as an observable adaptation resulting from accumulated changes over time (for example, stinging behavior in honeybees) (Suárez and Lloyd 2023; Marín 2024).

The units that interact with the environment, interactors (Hull 1980), are those that exhibit phenotypic variation, and whose interaction with the environment produces differences in replication or reproduction. By contrast, replicators or reproducers are the units of inheritance. The iteration of this process, requiring the joint action of interactors and replicators over time, gives rise to the emergence of adaptations (Marín 2024). From this perspective, natural selection can be defined as the process by which the differential proliferation of interactors causes the differential replication of replicators (or the differential reproduction of reproducers) (Suárez and Lloyd 2023).

The flow of information runs through the entire life process at its different levels. The more relevant environmental information an organism can capture, process, and use, whether acting as a replicator or as an interactor, the greater its effective fitness. Replicators interact among themselves within a single organism, but the organisms that contain them also interact with one another in many different ways (Dawkins 1999), generating symbiotic dynamics across multiple scales. Under such conditions, the amount of meaningful information increases, since each informational unit acquires greater functional significance as the system of which it is part becomes more interconnected and

coherent (Margalef 1996). All of this points toward an idea that has gained increasing support in recent years: life can be understood as matter that exchanges information and generates functional meaning, with biological evolution being a continuous process of information acquisition, transformation, and optimization (Davies 2016; Davies 2019; Adami 2024; Paredes et al. 2024). The different levels at which information and meaning are transmitted and generated can thus be viewed as a hierarchy of levels or categories of information handling.

However, since the beginnings of neo-Darwinism, evolution has often been equated with changes in gene frequencies, reducing the complexity of evolutionary processes to a purely genetic dimension (Fisher 1930; Wright 1931; Haldane 1932; Dobzhansky 1982). As we have seen, genes are the replicators, but the units that actually interact with the environment, those that experience selection, are the ones that exhibit phenotypic variation and differential proliferation.

Genetic variation is undoubtedly an essential component of the evolutionary process, but the reductionist tendency has been reinforced not only by the central role of genetic information but also by the pursuit of mathematical formalization of evolution. Such formalization becomes much simpler if the individual is represented as a combination of a few replicators, that is, as a simplified genotype. However, when the effect of a gene depends on a complex network of interactions, its real effect is a network property of the system in which the gene is embedded, together with its local environment (Lewontin 1974; Oyama 2000; Jablonka and Lamb 2006; Noble 2011; Marín and Wade 2025).

### **3. REPLICATORS, INTERACTORS, AND INFORMATION**

Information in organisms is not static. Although a gene possesses a fixed nucleotide sequence, its expression depends on context and on multiple internal and external conditions. Genes combine with other informational pathways, generating complex regulatory networks within and beyond the cell, which cooperate and interact until forming a coherent whole. Biological information is indeed encoded in genes (Adami 2004; Koonin 2016), yet these genes are expressed as proteins that participate in signaling networks and intercellular communication, as well as in interactions between

the cell and its environment. External signals are received through membrane receptors, triggering signaling cascades mediated by proteins such as kinases or G-proteins and transcription factors that ultimately regulate gene expression (Davies 2019; Azpeitia et al. 2020). Information in biology is therefore dynamic and functionally meaningful (Nurse 2008; Koonin 2016; Davies 2019; Cartwright et al. 2024). Living systems acquire information about their environment and integrate it into a hierarchical network of flows and transformations, very different from a linear sequence of zeros and ones. Moreover, information also travels between individuals, for example through species-specific recognition signals used in mating (Garcia et al. 2020).

From an evolutionary perspective, many approaches focus on how genomes and organisms store information about the environment to which they are adapted (Kimura 1961; Wagner 2007; Adami 2012; Baalen 2013; Wagner 2017). However, information is not accumulated solely in the genome; it can also be measured at the population level, from the informational increase caused by natural selection, by analyzing the distribution of frequencies within the population (Frank 2012) or even from the distribution of matings (Carvajal-Rodríguez 2018; Carvajal-Rodríguez 2024a). Consequently, although adaptive information accumulates in genomes throughout evolution, the resulting patterns can be measured at different levels: population, phenotypic, and genotypic (Hledík et al. 2022). At the population level, it is also possible to quantify the information associated with the different stages of the life cycle by comparing them with their corresponding null models (Smith 2024). From a quantitative standpoint, one possible metric to estimate adaptive information is the Kullback-Leibler divergence (KL) (Kullback 1997), which measures the deviation of an observed distribution from that expected under a neutral model. However, if the underlying dynamic model follows a replicator-type structure, either strict or extended to include mutation and recombination, so that a function  $\Phi$  maps the frequency change of entities from generation  $t$  to  $t + 1$  (Ay et al. 2017), then the frequency change associated with the logarithm of  $\Phi$  corresponds to the Jeffreys divergence (Frank 2012; Carvajal-Rodríguez 2024b). In this case, the Jeffreys divergence is the most appropriate metric for quantifying the total amount of information generated during the selective process, as it is a symmetric measure that incorporates changes in both directions.

Finally, at the ecosystem scale, species diversity and their distribution also constitute a quantifiable form of information, reflecting the structural and functional complexity of ecological systems (Margalef, R. 1968; Margalef 1996; Sherwin 2018; Konopiński 2020). Altogether, these levels, molecular, cellular, population, and ecosystemic; illustrate that life can be understood as a hierarchical network of informational flows.

In the following sections, we will show how these flows can be measured and subsequently reinterpreted as manifestations of Information Handlers (IH), whose transformations and replications generate information folds; i.e., cumulative, structured measures of informational change within the system, at each hierarchical level of organization.

### 3.1 Replicators

Let us briefly recall the model of replicator evolution. Consider a population of replicators, and let  $w_i(t)$  denote the fitness of type  $i$  in generation  $t$ , understood as its reproductive success potential or its capacity to generate copies of itself. The expected change in the frequency of type  $i$ , as a function of its relative fitness with respect to the population mean, is given by the discrete replicator equation:

$$p_i(t+1) = p_i(t) \frac{w_i(t)}{\bar{w}(t)} \quad (1)$$

where  $p_i(t)$  is the frequency of replicator  $i$  in generation  $t$ , and  $\bar{w}(t)$  is the mean fitness of the population in generation  $t$ . The change in frequency due to differential fitness is thus:

$$\Delta p_i = p_i(t+1) - p_i(t) = p_i(t)(\omega_i - 1)$$

where

$$\omega_i = \frac{w_i(t)}{\bar{w}(t)}$$

If, instead of considering the replicator  $i$ , we examine an associated phenotypic trait  $Z_i$ , the population mean value in generation  $t$  is

$$\bar{Z}(t) = \sum_{i=1}^n p_i(t) Z_i$$

and the mean value in the next generation, assuming that only frequencies change, is

$$\bar{Z}(t+1) = \sum_{i=1}^n p_i(t+1) Z_i$$

The change in the mean value of  $Z$  due to differential fitness can be expressed (Frank 2012) as

$$\Delta \bar{Z} = \text{Cov}(\omega, Z) \quad (2)$$

If we take the trait  $Z$  to be the logarithm of relative fitness,  $Z_i = \log(\omega_i)$ , then (2) corresponds to a divergence  $\Phi$  whose functional form is

$$\phi = J(p', p) = \sum_i (p'_i - p_i) \log \frac{p'_i}{p_i}$$

This is the Jeffreys divergence (Pardo 2018; Carvajal-Rodríguez 2024b), the symmetric version of the Kullback–Leibler divergence, which sums the two possible divergences before and after selection. In this case, (2) can be rewritten as

$$\Delta \bar{\log}(\omega) = \text{cov}(\omega, \log(\omega)) = J(p', p) \quad (3)$$

Equation (3) measures the rate of information production by natural selection (Frank 2012).  $J(p', p)$  quantifies how much information is gained by the fitter replicators and how much is lost by the less fit ones. An equation of the same form as (3) can also be obtained for extended replicators, for example, when including mutation or recombination; provided that there exists a function  $f(\omega)$  that maps a probability distribution  $p(t)$  into  $p(t+1)$ :

$$p_i(t+1) = p_i(t) f(\omega_i) \quad (4)$$

and allows defining  $Z = \log(f(\omega))$  as an informational trait. In this way, the evolutionary dynamics can be described in terms of the Jeffreys divergence, which measures the total flow of information reallocated during the life-cycle process.

Carvajal-Rodríguez (2018; 2019; 2020; 2024a; 2024b) applied this framework to different contexts: non-random mating, sexual selection, and haplotype evolution; showing that the deviation from the corresponding null model is naturally expressed through  $J(p', p)$ . In all these cases, the term  $J$  represents a symmetric measure of the total change in information, capturing the net informational flow in the population.

### 3.2 Interactors

As we have seen, the units of selection can be located at different hierarchical levels: replicators, which transmit information, and interactors (or vehicles), which express that information and act upon the environment. This distinction helps to disambiguate the different roles biological entities play in evolution. A single entity can behave as a replicator in one context and as an interactor in another, depending on the level of organization under consideration (Suárez and Lloyd 2023). This raises the question: *can the change in the frequency distribution of a population of organisms, or other types of interactors, be described by equations of type (1-4)?* We know that, at least in the case of mating interactions, the answer is yes. Carvajal-Rodríguez (2018) showed that mating distributions can be represented by an equation of the same form as (1), where the distributions correspond to the frequencies of pair formation under different schemes (random, mate choice, or intra-sexual competition). Although mating pairs are not replicators in the strict sense, since they do not copy themselves, the dynamics of pair formation described by the equation generate measurable information in terms of the divergence  $J$ . In this context,  $J$  quantifies the total information associated with the change in the distribution of matings, such that when mate choice deviates from randomness, detectable information is generated both in the mating distribution and, potentially, in the general population (when sexual selection is present).

In a similar way, interactors can also be defined at higher levels: groups, symbiotic associations, or ecological communities. The question that then arises is: *what kind of information is generated at the population or supra-individual level in these cases?* This question motivates the next two sections, which show how changes in frequency distributions within or between groups, or in holobionts, can

be described by equations of type (1-4) and how the associated information can be quantified through the measure  $J$ .

#### 4. INTRA- AND INTER-GROUP SELECTION INFORMATION

In the context of group selection, natural selection can act at different hierarchical levels that often exert forces in opposing direction (Okasha 2008; Frank 2024). Frank (2025) develops a simple tragedy of the commons model (Hardin 1968) to describe individual competition at the local level and group competition at the global level between two types of haploid genotypes with competitiveness values  $x_1$  and  $x_2$ , occurring at frequencies  $q$  and  $1 - q$ , respectively.

##### **Intra-group information**

The mean competitiveness within a group is

$$y = q x_1 + (1 - q) x_2$$

and the relative competitiveness of genotype 1 is

$$\omega_1 = \frac{x_1}{y}$$

Its frequency within the group after intragroup competition is

$$q' = q \omega_1$$

The change in within-group frequency before between-group selection is

$$\Delta q = q' - q = q \omega_1 - q = q(\omega_1 - 1),$$

and the average change in an associated character  $Z$  can be written in the form of an equation of type (2). If we take  $Z = \log(\omega)$ , we obtain the within-group information

$$J_q = (q' - q) \log \frac{q'}{q} + (q - q') \log \frac{1 - q'}{1 - q}$$

which represents information generated by selection within a particular group.

The average intra-group information before inter-group selection will be

$$\bar{J}_x = \sum_q p_q J_q$$

where  $p_q$  is the probability that the group has an initial frequency  $q$ .

### Inter-group information

The relative competitiveness of the entire group (normalized by the average group fitness) with a relative intensity  $s$  of inter-group selection is (eq. 11 in Frank 2025)

$$\omega_G = \frac{(k - y)^s}{\bar{w}},$$

where  $k > 0$  is a term that avoids negative values in  $\omega_G$ .

The probability that a group with genotype 1 at frequency  $q$  contributes to the set of groups of the next generation is proportional to

$$p'_q = p_q \omega_G$$

Information on the change in group distribution caused by inter-group selection, taking  $Z = \log(\omega_G)$ , is

$$J_G = \sum_q (p'_q - p_q) \log \frac{p'_q}{p_q}$$

Thus, for a simple tragedy of the commons model, it is possible to quantify the flow of information associated with intra- and inter-group selection based on changes in frequency distributions.

However, real organisms are part of complex interaction networks where, in addition to competition between groups, individuals of different species are involved, and their cooperation may be essential for survival. This is particularly evident in symbiotic processes, which leads to a new question: *what kind of information is generated in the combination of organisms that form part of symbioses?*

## 5. EXTENDED PHENOTYPE AND SYMBIOSIS

In *The Extended Phenotype* (Dawkins 1999), Dawkins proposed that the effects of a gene, that is, of replicators, are not confined to the phenotype of their survival machine (the individual or interactor sensu Hull 1980), but can extend beyond it, significantly influencing the environment. The nests built

by birds, the dams of beavers, or the webs of spiders can be considered expressions of these animals' genes, since such constructions and behaviors directly affect the organism's survival and reproduction.

Natural selection acts upon these extended phenotypes in the same way it acts upon traditional phenotypes: genetic variants responsible for extended phenotypes that produce favorable effects for the individuals carrying them will tend to be positively selected. Extended phenotypes associated with a gene or set of genes—such as nests, burrows, or river dams—generate effects that transcend the individual, modifying the dynamics of entire groups. Thus, the extended phenotype can be interpreted as a mechanism through which the effects of genes at one level (the individual) influence selection at another level (the group). Dawkins also recognized that mutually beneficial replicators tend to predominate over others. From his gene-centered perspective, this explains the evolution of cooperative vehicles that carry replicators with complementary effects. However, when an organism, through a specific genetic variant, modifies the environment or the behavior of another organism in such a way that both benefit, the result is functionally analogous to a mutualistic symbiosis. A classic example is that of intestinal bacteria that help digest food in exchange for a stable habitat (Edelaar and Bolnick 2019); bacterial genes that favor this relationship spread because they increase their own survival and reproduction (Bosch et al. 2024).

While Dawkins offered a replicator-centered explanation of cooperation, Margulis (Margulis 1976; Margulis and Bermudes 1985) approached symbiosis from an ecological and organizational perspective, showing how such interactions can give rise to new levels of biological complexity. From this standpoint, symbiosis is not merely an adaptive strategy but also a process that generates organization and facilitates the exchange of information between systems. When we refer specifically to the symbiosis between a host and its microbial communities, we are talking about holobionts, that is, integrated systems that constitute new functional and evolutionary units.

## 5.1 Holobionts

A holobiont is a composite organism that includes a host together with the ecological community of microorganisms that constitute its microbiome (Margulis 1991; Roughgarden 2020). The combined set of genomes of the host and its associated microorganisms forms the hologenome (Zilberman-Rosenberg and Rosenberg 2008; Theis et al. 2016). The hologenotype represents the particular configuration of the hologenome in an individual holobiont, and its phenotypic expression, which determines the behavior, physiology, and morphology of the system, defines the holophenotype. Holobiont selection is therefore understood as the differential reproduction or survival of holobionts according to their holophenotypes (Roughgarden 2020). Over time, this selection modifies both the total number of holobionts and the frequency distribution of hologenotypes.

A current debate concerns whether selection acts on the host-microbiome combination as a single functional unit, or whether the two coevolve partially independently, given that microbiome transmission is not always vertical (Roughgarden 2020 and references therein). In many cases, hologenome integration is incomplete and microbiome transmission is horizontal, which makes it difficult to regard the holobiont as a fully autonomous unit of selection. Nevertheless, a model proposed by Roughgarden (2020), shows theoretically that holobiont selection remains an evolutionarily plausible force even under horizontal transmission. The model includes three stages within each generation: (1) microbial transfer among holobionts, (2) internal microbial proliferation, and (3) holobiont selection.

When transmission is horizontal, the first step is replaced by transfer from external microbial pools.

If  $H''(t, n)$  is the relative frequency of holobionts containing  $n$  microbes after microbial proliferation, and  $W(n)$  is the fitness of a holobiont with  $n$  microbes, the frequency of holobionts with  $n$  microbes after reproduction is expressed as (c.f. equation 9 in Roughgarden 2020)

$$H(t+1, n) = \frac{H''(t, n)W(n)}{\sum_n H''(t, n)W(n)}.$$

This equation predicts the frequency of the different classes of holobionts at the beginning of the next generation and constitutes a replicator-type equation. Within this framework, the rate of information production due to holobiont selection corresponds to the Jeffreys divergence  $J$  between the distributions before and after reproduction.

Using a different approach, (Week et al. 2025) apply a niche-construction model from quantitative genetics, defining a host trait  $Z$  determined additively by a genetic component  $G$  and a microbial component  $M$  ( $Z = G + M$ ). If  $W$  denotes the fitness of the host, the effect of selection on the trait is expressed as

$$\text{cov}(W, Z) = \text{cov}(W, G+M) = \text{cov}(W, G) + \text{cov}(W, M)$$

so that selection acting on  $Z$  results in indirect selection on both the host genotype and its microbiome (Week et al. 2025).

If we take  $Z = \log(W)$  and assume that fitness decomposes into three multiplicative components,  $W_G$  associated with the host's genetic component,  $W_M$  associated with the microbial component, and an interaction term  $a_{GM}$  defining the affinity or association between  $G$  and  $M$ :

$$W = W_G W_M a_{GM}, \quad \log W = \log W_G + \log W_M + \log a_{GM},$$

then the Jeffreys divergence associated with the mean informational change in  $\log(W)$  can be expressed (see Mathematical Appendix 9.1 for details) as

$$J = J_G + J_M + I_{assoc} \quad (5)$$

where  $J_G$  and  $J_M$  measure the total information associated with the frequency changes mediated by selection on the genome and on the microbiome, respectively, and  $I_{assoc}$ , measures how much information the system uses in pairing particular genotypes with particular microbiotas. In a certain sense, this last term constitutes the signature of stable symbiotic coevolution.

From equation (5), it would be possible to design statistical tests and estimators analogous to those developed in (Carvajal-Rodríguez 2018; Carvajal-Rodríguez 2020) that, using genetic frequencies in  $G$

and M, could test for selection acting on G and/or M, as well as for possible associations between host and microbial genomes.

These results connect directly with the next step of this work: to formalize the informational dynamics of Information Handlers (IH) within categorical hierarchies of replication and interaction.

## 6 EVOLUTION AS CATEGORIES OF INFORMATION

We have seen that in the evolution of life, everything appears interconnected, genes, individuals, groups, symbionts, and environment; while at the same time being profoundly differentiated and dynamic. Life manifests as a fluid network of organized interactions. Within this view, the concept of relational biology takes on its full meaning. Relational biology (Rosen 1991), inspired by the ideas of Rashevsky (1954), seeks to understand living systems not through their material components, but through the organizational and causal relations that constitute them. Its goal is not to describe what a living being is made of, but rather how it is organized and how it maintains the relationships that make it alive. Rosen proposed that organisms are systems closed to efficient causation, that is, all the causes that produce and sustain the system are contained within it. In other words, the system produces both its own parts and the rules that govern its functioning. To formalize this idea, Rosen used category theory (Mac Lane 1971; Cheng 2022; Leinster 2025), representing organisms as (M,R) systems, where M corresponds to metabolic processes (transformations of matter and energy), and R to repair or reproductive processes that generate the very components of M. The (M,R) system is therefore causally closed: the functions that maintain it are produced by the system itself without the need for external causes.

Unlike the reductionist and gene-centered approach, relational biology, while not denying the importance of material components or mechanisms, emphasizes that life is not explained by its parts, but by the functional relations among them (Lane 2024). The laws of life are laws of organization, not of matter. Life would thus be an organizational category, not a material one. This approach connects with other systemic or holistic frameworks, such as autopoiesis theory (Varela et al. 1974; Maturana and Varela 2005), organizational closure (Mossio and Moreno 2010) and biosemiotics

(Pattee 2001). A cell can be understood as a molecular realization of an (M,R) system (Rosen 1972), and autopoietic systems can be considered a subset of (M,R) systems (Letelier et al. 2003). Within this framework, each level of organization constitutes a complete (M,R) system, and simultaneously a component of a higher-order (M,R) system. Relational closure is therefore nested, not isolated: an isolated cell loses its causal network because its boundary conditions disappear, but its functional autonomy persists within a broader relational context. Autonomy does not imply physical independence, but rather functional coherence within a larger system.

This raises the question of how to incorporate the informational and sensory openness of organisms into a framework that postulates causal closure. Even though exchanges with the environment, of matter, energy, or signals, do occur, these exchanges do not determine the system's internal organization, since the system interacts with the environment but is not defined by it. What remains closed is the internal causal network that generates and maintains its own functions. In this context, information is understood in a relational, not transmissive, sense, not as a thing, but as a relationship between things. In the (M,R) model, the arrows (functors in categorical terminology) that connect one level to another, transport structures, not matter, i.e., they preserve correspondences. This transport can be interpreted as a flow of semantic information, if we understand information as the preservation of structure. What is essential is the meaning that such information acquires for the different organic entities; that meaning, expressed in (M,R) relations or in equivalent semiotic models, defines the organism itself.

Nevertheless, organization implies a quantifiable informational substrate, even if its meaning cannot be reduced to it. Physical information may be necessary but not sufficient to explain living systems. Every organized system entails a quantifiable informational pattern, but that measure does not capture the organizational meaning that determines its functional role within a causally closed system (Rosen 1991; Pattee 2001; Hoffmeyer 2008). To integrate the physical notion of information with the organizational and semiotic notions of meaning and function, the multilevel information paradigm emerges. This paradigm holds that living systems do not handle information at a single level, molecular or genetic, but across multiple hierarchical levels. Each level (molecular, cellular,

physiological, ecological) generates, transmits, and interprets information within its own domain, and these levels interact hierarchically, integrating information both upward and downward (Walker and Davies 2013; Davies 2019). In this sense, life can be understood as a multilevel integration of information, where the lower levels contribute variability and the higher levels contribute organization and meaning. This relational and hierarchical view of information forms the basis upon which we can formalize evolution as a categorical network of information: the Information Handler (IH) model.

## **6.1 Categorical IH Model: Information Handlers with (M,R) Closure**

What follows outlines a categorical model that represents an attempt to apply category theory to a population dynamics framework of genetic type, such as an RNA world (self-replicating entities with rudimentary metabolism), while also integrating some of the principles of relational biology. There already exist more elaborate developments that aim to express genetic notions and processes within a categorical framework. For example, Tuyéras (2018) describes structures such as DNA sequences, haplotypes, linkage, and recombination in categorical terms, seeking to capture the details of biological processes, and in later work connects them to, for instance, the identification of haplotypes responsible for specific phenotypes in a population or the conceptualization of population stratification structures applied to patient data (Tuyéras 2023). Given the complexity of genetic frameworks that articulate genotype–phenotype relations, the resulting formalisms are necessarily extensive and sophisticated. There are also other attempts to use category theory in genetics and evolution, for example, to capture the structure and dynamics of phylogenetic trees (Baez and Otter 2017) or to model evolutionary dynamics such as Moran processes (Mahadevan 2024). Likewise, Wu (2023) proposes employing category theory to provide a more holistic and structured view of the genome, beginning by defining a category of the gene, whose morphisms represent gene positions, interactions between genes, or relationships in genetic ontology.

The model presented here is conceptually much simpler than those mentioned above. It uses only the minimal categorical structure required, without attempting to describe genetic phenomena in

detail, and thus does not need to formalize all the underlying complexity. An abstract entity called an Information Handler (IH) is defined, endowed with self-replicative, mutational, and recombinational capacities, and the rules that internally generate those functions are formalized so that the system becomes closed. The IH model seeks to connect the reductionist approach (focused on genes and fitness) with the holistic approach (which prioritizes interactions and internal organization) within a flexible formal framework. The model describes the evolutionary dynamics of a biological system composed of entities that handle information in ways that enable them to replicate and survive transformations, whether internal or environmental. Each entity (IH) represents an informational processor capable of copying, varying, and reproducing itself, thus generating a structure analogous to an RNA world, but expressed in categorical and informational terms.

The central idea is that biological organization is conceived as a system of processes that generate and maintain themselves, such that the functions producing the system's transformations are also generated internally. This principle is expressed formally through the closure morphisms ( $M, R$ ). At the same time, biological organization can be viewed as a dynamic process that produces information. Thus, evolutionary dynamics are translated into informational folds, measured by Jeffreys divergences, which quantify the information gained through selection, transmission, and reproduction. A summary of the model's key features, its categorical structure, and its informational meaning is presented below. The complete formal model is given in Mathematical Appendix 9.2.

### *6.1.1 Categorical Structure*

A category consists of objects and morphisms (arrows representing transformations). An initial object  $X(0)$  is defined. From  $X(0)$ , variants  $A(1), A(2), \dots, A(t); B(1), B(2), \dots, B(t); \dots$  are generated through replication or reproduction (see below), with or without mutation, where  $t$  represents discrete generations from the initial configuration. For each lineage  $\alpha \in \{A, B, C, \dots\}$ : at each step  $t$ , the classes  $\alpha_0(t), \alpha_1(t), \dots, \alpha_s(t)$ , coexist, representing the accumulated substates of the lineage. Operationally, each lineage  $R_\alpha$  is treated as a small category whose objects are the possible states of the lineage and whose elementary morphisms  $f_{\alpha,t}$  describe discrete temporal transitions from step  $t$  to  $t+1$ .

### *Objects*

The objects correspond to the different states within the lineage:

$$Ob(R_\alpha) = \{\alpha(0), \alpha(1), \dots, \alpha(t)\}$$

Each object represents a type or subtype of the Information Handler (IH) with certain accumulated variations relative to the ancestor. Thus, at step  $t > 0$ , the possible subtypes within lineage  $\alpha$  are

$$\{\alpha_0(t), \alpha_1(t), \dots, \alpha_s(t), \dots, \alpha_t(t)\}$$

where the subscript may indicate the number of accumulated variations relative to the origin or any other characteristic defining the substate. Hence,  $\alpha_0$  corresponds to the ancestral type.

### *Morphisms (arrows)*

Morphisms describe the operations of replication, variation, or transformation that lead from one state to the next:

$$f_{\alpha,t}: \alpha(t) \rightarrow \alpha(t+1)$$

where  $\alpha(t)$  is the Information Handler at step  $t$ , and  $f_{\alpha,t}$  is the combined action of replication, with or without mutation, producing the new handler  $\alpha(t+1)$ .

At this level, it is not necessary to specify internal subtypes, since the arrow refers to the entire set or distribution of the lineage at time  $t$ . As we will see, each internal state  $s$  of the handler  $\alpha_s(t)$  is associated with a replication rate  $r_{as}(t)$ .

The morphism  $f_{\alpha,t}$  represents the discrete temporal transition from step  $t$  to  $t+1$ . It may or may not involve a change in the internal substate  $s$  depending whether replication occurs without mutation ( $s' = s$ ), preserving the internal configuration, or with mutation or variation ( $s' \neq s$ ), producing a new subtype within the lineage. This morphism thus captures the causal process by which the system generates its next configuration, either identical or modified, maintaining lineage continuity while potentially introducing informational change.

### *Closure*

Each lineage category possesses an internal (M, R) closure:

$$(M, R): \alpha(t) \xrightarrow{\rho_t} [\alpha(t), \alpha(t+1)] \xrightarrow{ev_t} \alpha(t+1)$$

such that

$$ev_t \circ (\rho_t \times id_{\alpha(t)}) = f_t$$

where  $\rho_t$  assigns to each state the mechanism capable of reproducing it in the future, an abstraction of the production of enzymes or functions enabling replication, and  $ev_t$  applies the internally generated transformations to perform the transition to the next state (see Appendix 9.2.3.1).

### *Supercategory R*

The collection of lineages forms a category  $R$ , whose objects are the categories  $R_\alpha$  and whose morphisms represent processes of reproduction or interaction between lineages (e.g., hybridization, gene transfer, symbiosis, etc.):

$$R = \{R_\alpha\}_{\alpha \in \Lambda}$$

$$Mor(R) = \{F_{\alpha \rightarrow \beta}: R_\alpha \rightarrow R_\beta\}.$$

#### *6.1.2 Discrete Dynamics within a Lineage*

Each object at substate  $s$  in lineage  $\alpha$  is connected to the set of copies it contains i.e. it has a population of instances

$$P_\alpha(\alpha(s); t) \subseteq Set$$

with cardinality  $N_{\alpha s}(t)$  and the frequency of subtype  $s$  at time  $t$  is

$$p_{\alpha, s}(t) = \frac{N_{\alpha, s}(t)}{\sum_u N_{\alpha, u}(t)}.$$

The evolution of frequencies follows the replicator equation:

$$p_{\alpha, s}(t+1) = \frac{p_{\alpha, s}(t) \cdot r_{\alpha, s}(t)}{\sum_u p_{\alpha, u}(t) \cdot r_{\alpha, u}(t)}, \quad (6)$$

where  $r_{\alpha, s}(t)$  is the effective replication rate including both the effects of selection and transmission.

However, it is also possible to consider the strict replication rate to measure frequency change due

solely to selection (see Sections 9.2.3–9.2.4 of the Mathematical Appendix). This is the same discrete replicator equation introduced in Section 3, but now with lineage ( $\alpha$ ) and subtype ( $s$ ) indices.

Within each generation  $t$ , older (persistent) and new states coexist simultaneously. The change in the frequency distribution from  $t$  to  $t+1$ , according to equation (6), allows us to define a replicative informational fold measured by the Jeffreys divergence (see below):

$$\text{Info}_{\alpha}^{(rt)}(f_{\alpha,t}) = J(p_{\alpha}(t+1), p_{\alpha}(t)).$$

### 6.1.3 Reproduction

In addition to replication within lineages, the model allows for reproduction or recombination between IH entities (Figure 1). The reproductive morphisms are represented as pairings modeled by a binary operation (reproductive bifunctor) that takes two lineages and produces joint offspring. That is, each possible pairing between a pair of IHs from lineages  $\alpha$  and  $\beta$  has a mating fitness  $m_{\alpha(t),\beta(t)}$ , and the resulting pairing frequencies are  $q'_{\alpha(t),\beta(t)} = q_{\alpha(t),\beta(t)} m_{\alpha(t),\beta(t)}$ , which can be compared with those expected by chance  $q_{\alpha(t),\beta(t)}$  to define a reproductive informational fold measured by the Jeffreys divergence:

$$J_{rep}(t) = J(q'(t), q(t)) = \sum_{i,j \in \alpha, \beta} (q'_{ij}(t) - q_{ij}(t)) \ln \frac{q'_{ij}(t)}{q_{ij}(t)}.$$

The mating distribution can be decomposed into intra-lineage ( $\alpha = \beta$ ) and inter-lineage ( $\alpha \neq \beta$ ) components, and the corresponding informational folds can be calculated. This allows, for instance, modeling and analyzing sexual isolation between lineages, negative assortative mating, and similar processes (see Section 9.2.6 of the Mathematical Appendix for details).

After mating, the effective replication rates are determined by a natural transformation that combines rates:

$$\phi_{\alpha,\beta}: (r_{\alpha} \times r_{\beta}) \rightarrow (r_{\alpha\beta}^{\text{post}} \circ \hat{m}_{\alpha,\beta})$$

whose components define the aggregating rule

$$r_{ij}^{\text{post}}(t) = G(r_{\alpha,i}(t), r_{\beta,j}(t))$$

which determines the fecundity associated with the cross  $i \times j$ . Different choices of the function  $G$  correspond to different hypotheses about fecundity following mating. For example:

$$G_L(x, y) = x, G_R(x, y) = y, G_{mean}(x, y) = (x + y)/2,$$

$$G_{geom}(x, y) = \sqrt{xy}, G_\lambda(x, y) = \lambda x + (1 - \lambda),$$

$$G_{hybrid}(x, y) = I_{xy}(x + y)/2$$

where  $I_{xy} \in [0, +\infty)$  is a sexual isolation factor.

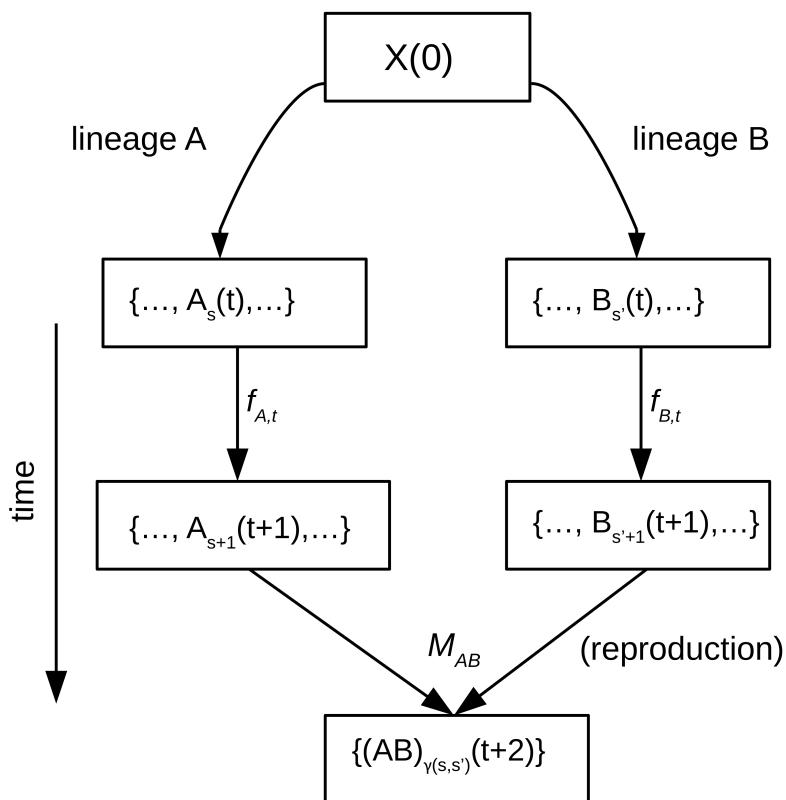


Figure 1. Basic directed graph of the IH model. The initial state  $X(0)$  gives rise to lineages A and B, which evolve in discrete time steps  $t$ ; the arrows indicate the intra- and inter-lineage transformations.

#### 6.1.4 Information Functor and Biological Interpretation

In category theory, a functor is a rule that “translates” one category into another while preserving its basic structure. Each object in the first category (for instance, a mutational state or a population) is assigned to an object in the second category (for example, its statistical or informational description), and each arrow between objects in the first category (for example, replication) is mapped to an

arrow between the corresponding objects in the second category (the associated informational change).

In our IH model, we have two interconnected levels: (1) a biological level, where material transformations occur, reproduction/recombination, replication, and mutation, so that Information Handlers (IHs) combine, copy, and change; and (2) an informational level, where we measure the effects of those transformations in informational terms, i.e., how the distribution of types within a lineage changes, how much new information appears, and how much information is associated with intra- and inter-lineage pairing events, among others.

The information functor acts as a bridge between these two levels, the biological and the informational. It translates between the world where IHs reproduce, copy, and mutate, and the informational world where these processes manifest as changes in frequencies and in the system's total information content.

The informational change within each lineage is measured by the following information functor:

$$\text{Info}_\alpha(f_{\alpha,t}) = J(p_\alpha(t+1), p_\alpha(t)),$$

where  $J(p_\alpha(t+1), p_\alpha(t))$  is the Jeffreys divergence, quantifying the amount of information gained by moving between the distributions  $p_\alpha(t)$  and  $p_\alpha(t+1)$ . Two variants are defined,  $J_r$ , which measures the informational change caused by selection, and  $J_{rt}$  which measures the change caused by selection plus transmission (mutation, recombination, and migration). Biologically,  $J_r$  measures how much of the informational gain is due to selection acting on the differential fitness of existing types, while the difference  $J_{rt} - J_r$  measures the portion attributable to non-strictly selective processes that redistribute information without conferring an immediate fitness advantage (see details of this partition in Section 9.2.5 of the Mathematical Appendix).

#### 6.1.5 Total Informational Fold

The fold, or total informational summary, generated by the system at step  $t$  represents the accumulated quantity of informational change due to replication, transmission and reproduction:

$$\text{Info}_{Total}(t) = J_{repr}(q'(t), q(t)) + \sum_{\alpha \in \Lambda} J_{rt}(p_\alpha(t+1), p_\alpha(t))$$

This total information fold functions as an informational closure, in the sense that it ensures each step of the biological dynamics has a corresponding informational counterpart.

The total information functor  $\text{Info}_{\text{Total}}$  assigns to each biological morphism  $f_{a,t}$  its corresponding informational measure  $\text{Info}_{\text{Total}}(t)$ , quantified by the divergences  $J$ . The global accumulated information over a discrete interval  $[0,T]$  is then obtained as the sum of the information measures assigned by the functor to each morphism over time:

$$\text{Info}_{\text{Global}}^{\text{Total}}(T) = \sum_{t=0}^{T-1} \text{Info}_{\text{Total}}(t).$$

## 7. DISCUSSION

In this work, we have briefly reviewed several representative evolutionary models, from the simple replicator equation (Wright 1931; Dawkins 1978; Bürger 2000), to the tragedy of the commons model with intra- and intergroup competition (Frank 2025), and more recent models of holobiont selection (Roughgarden 2020; Week et al. 2025), to show that evolutionary change in the distribution of frequencies can be expressed through the Jeffreys divergence (Frank 2012). The same divergence also captures the redistribution of pairings caused by differential mating fitness (Carvajal-Rodríguez 2018; Carvajal-Rodríguez 2024a).

Building on these foundations, we have outlined a formal conceptual framework based on category theory and centered on the concept of the Information Handler (IH). Information Handlers are entities endowed with the capacity for self-maintenance and replication, but also for mutation and combination with other handlers to generate mixed copies. In our scheme, we start from an original handler from which different lineages develop on a discrete temporal scale. Within each lineage, we distinguish different mutational states, implying distinct replicative and reproductive capacities. Each lineage is, in turn, an object of a higher-order category in which the relations between lineages may correspond to processes such as hybridization, sexual isolation, or symbiosis. Category theory allows us to define these objects and relationships in an abstract way, so that a lineage of handlers could equally represent ribozymes, bacteria, animals, or plants.

At each process, replication (with or without selection or mutation), or intra- or inter-lineage reproduction, we define an informational fold that captures the corresponding change in the distribution of entities involved at each step. The process is closed: the arrows or relations that carry a handler to its copy or descendant are generated within the system itself. This scheme integrates aspects of a reductionist or gene-centric approach with those of a systemic or organizational perspective. This is particularly relevant because, in contemporary studies of life and evolution, two seemingly opposing schools coexist: on one hand, the classical neo-Darwinian approach, reductionist and gene-centered; and on the other, a systemic or organizational approach that emphasizes the biology of organization and tends to relativize the central role of genetics and evolution in explaining life (Mossio and Pontarotti 2019; Noble and Noble 2023; Lane 2024; Mossio 2024; Vega 2024; Noble and Noble 2025).

Rosen's (1991) proposal of relational biology has been revisited by several authors who have reformulated the concept of organizational closure to explain self-determination and self-control in living systems (Montévil and Mossio 2015). According to this reformulation, biological organization can be understood as a closure of constraints, in which organisms are thermodynamically open systems (because of energy and matter flows) but organizationally closed at a distinct causal level. Constraints are relatively stable structures or dynamics that channel those flows, such as enzymes, membranes, or blood vessels, and whose existence depends on other processes that they themselves enable. Under this view, life is characterized by an organization that maintains itself through a network of mutually dependent constraints. This closure of constraints generates self-control and stability without eliminating variability: the system preserves its organization while allowing functional and evolutionary change.

Although some formulations of relational or organizational biology do not explicitly refer to information, other approaches do attempt to link information, meaning, and causation (Walker and Davies 2013; Barbieri 2019; Prinz 2023; Yurchenko 2023). In any case, the concept of information remains central in biology (Davies 2016; Davies 2019; Avery 2021; Adami 2024; Paredes et al. 2024), from replicators, genes, and haplotypes that store information about an organism's development

and behavior, to the information associated with cellular communication signals, or, at the interactor level, with mating behavior, communication among individuals, and environmental perception. Biological information is transmitted, interpreted, and reorganized across multiple levels, including the population level. As we have seen, the distribution of matings in a population contains more information, released as an observable pattern, if pairs differ in mating fitness, and less when it is more random. Similarly, the distribution of haplotype frequencies in a population contains more information if genotypes differ in fitness, and less if they are equivalent.

However, the use of the concept of information in biology has been diverse and sometimes criticized as a mere metaphor (Griffiths 2001) though perhaps a useful one (García-Sancho 2011; Anta and Sánchez-Dorado 2025). Metaphor or not, as Davies (2019), emphasizes, in biology the amount of information is less relevant than its effect within the system that interprets it. Syntactic information may be seen as material, while semantic information is form; both are necessary, but only their coherent integration produces life or cognition. Information is understood as an emergent relational property, not as an independent substance or cause. Informational constraints (patterns, codes, structures) modulate local physical laws without violating them, for example, DNA constrains which proteins can form, and a neural circuit channels electric currents into a functional pattern. Information emerges when material dynamics reach a certain degree of self-constraint or organization.

The framework outlined in this first work, which is part of a larger project still in its early stages, serves firstly to highlight how the effect of natural selection can be captured from the corresponding changes in frequency distributions, whether in individuals, pairs, groups or holobionts; and secondly, to illustrate how the formalism of category theory can provide a powerful tool for defining integrative conceptual frameworks capable of encompassing the complex and diverse reality of biological and evolutionary processes. Within this theoretical framework, it is suggested that life can be understood as a process of managing, transforming, and closing information, where "closure" implies that each informational step is generated through internal morphisms. While information physically enters and leaves, its interpretation and causal relevance are internal to the system. There

is no external informational input that determines the transitions; changes depend on closure morphisms (M,R). Informational closure does not imply isolation from the environment, but semantic autonomy: living systems generate and maintain internally the informational transformations relevant to their organization. Thus, while life is thermodynamically open, it is informationally closed in the sense that its processes of encoding, interpretation, and transmission are sustained within a self-consistent causal network. This endogenous closure grants organisms their functional and evolutionary coherence. Life, therefore, may be viewed as a causally organized flow in which structure, function, and meaning emerge together.

## 8. CONCLUSION

The proposed framework connects gene-centered evolutionary dynamics—where genetic variants affect the fitness of organisms and their descendants—with relational and organizational perspectives that describe life through hierarchical levels of information processing. It suggests that biological organization can be represented as a causally closed network of informational processes, in which each transformation and its functional meaning are generated internally. In this view, evolutionary and organizational dynamics can be recaptured as informational divergences across different levels, providing a unified categorical framework that may help connect the reductionist genetic approach with relational biology.

## 9 MATHEMATICAL APPENDIX

### 9.1 SELECTION INFORMATION MODEL IN THE HOLOBIONT

Before selection

$$q_{ij} = p_i m_j$$

#### 9.1.1 Multiplicative selection with affinities

$$W_{ij} := W_G(i) \times W_M(j) \times a_{ij}$$

$$\Omega := \sum_{i,j} p_i m_j W_{ij}$$

$$q'_{ij} = p_i m_j \frac{W_{ij}}{\Omega}$$

### 9.1.2 Marginal frequencies after selection

$$p'_i = \sum_j q'_{ij} = p_i W_G(i) \frac{A_i}{\Omega}$$

$$m'_j = \sum_i q'_{ij} = m_j W_M(j) \frac{B_j}{\Omega}$$

where

$$A_i := \sum_j m_j W_M(j) a_{ij}$$

$$B_j := \sum_i p_i W_G(i) a_{ij}$$

and the residual affinity with respect to the new marginal frequencies:

$$a'_{ij} := \frac{q'_{ij}}{p'_i m'_j} = a_{ij} \frac{\Omega}{A_i B_j}.$$

because

$$q'_{ij} = p_i m_j \frac{W_{ij}}{\Omega} = p'_i \cdot m'_j \cdot a'_{ij} \Rightarrow p_i m_j \frac{W_{ij}}{\Omega} = \frac{p_i W_G(i) A_i}{\Omega} \cdot \frac{m_j W_M(j) B_j}{\Omega} \cdot a'_{ij} \Rightarrow a_{ij}/\Omega = \frac{A_i B_j}{\Omega^2} a'_{ij}$$

so

$$a'_{ij} = \frac{a_{ij} \cdot \Omega}{A_i B_j}$$

### 9.1.3 General information and partitioning

Taking  $Z = \log(W/\Omega)$  the average change of  $Z$  due to selection is

$$J := \sum_{i,j} (q'_{ij} - q_{ij}) \log \frac{q'_{ij}}{q_{ij}}$$

and we obtain the partition

$$J = J_G + J_M + I_{assoc}$$

where

$$J_G = \sum_i (p'_i - p_i) \log \frac{p'_i}{p_i} = \sum_i (p'_i - p_i) \log \frac{W_G(i) A_i}{\Omega}$$

$$J_M = \sum_j (m'_j - m_j) \log \frac{m'_j}{m_j} = \sum_j (m'_j - m_j) \log \frac{W_M(j) B_j}{\Omega}$$

$$I_{assoc} := \sum_{i,j} (q'_{ij} - q_{ij}) \log (a'_{ij})$$

#### 9.1.4 Verification

$$q_{ij} = p_i m_j, \quad q'_{ij} = p'_i m'_j a'_{ij}$$

$$\log \frac{q'_{ij}}{q_{ij}} = \log \frac{p'_i}{p_i} + \log \frac{m'_j}{m_j} + \log (a'_{ij}).$$

Then we can express  $J$  as

$$J = \sum_{i,j} (q'_{ij} - q_{ij}) \cdot [\log \frac{p'_i}{p_i} + \log \frac{m'_j}{m_j} + \log (a'_{ij})]$$

We separate it into three sums and note that

$$\sum_j (q'_{ij} - q_{ij}) = p'_i - p_i$$

and

$$\sum_i (q'_{ij} - q_{ij}) = m'_j - m_j,$$

we obtain

$$J_G := \sum_i (p'_i - p_i) \log \frac{p'_i}{p_i}$$

$$J_M := \sum_j (m'_j - m_j) \log \frac{m'_j}{m_j}$$

$$I_{assoc} := \sum_{i,j} (q'_{ij} - q_{ij}) \log (a'_{ij}) = \sum_{i,j} (q'_{ij} - q_{ij}) \log \frac{a_{ij} \cdot \Omega}{A_i \cdot B_j}$$

Multiplicative selection moves the marginals ( $J_G$ ,  $J_M$ ) through  $W_G$  and  $W_M$ , while the host-microbiota pairing structure is captured in  $I_{assoc}$  via affinity  $a_{ij}$  versus their row/column averages ( $A_i$ ,  $B_j$ ).

## 9.2. CATEGORICAL MODEL OF "INFORMATION HANDLERS" WITH CLOSURE (M,R) AND INFORMATION (INFORMATION FOLD)

### 9.2.1 Quiver and Lineage Categories $R_\alpha$ (IH)

#### 9.2.1.1 Basic Quiver (Directed Graph)

An initial object  $X(0)$  is defined. From  $X(0)$ , variants  $A(1), A(2), \dots, A(t); B(1), B(2), \dots, B(t); \dots$  are generated through replication with or without mutation, where  $t$  represents discrete generations from the initial configuration. For each lineage  $\alpha \in \{A, B, C, \dots\}$ : at each step  $t$  the classes  $\alpha_0(t), \alpha_1(t), \dots, \alpha_t(t)$ , coexist, representing the accumulated substates of the lineage. This step-by-step dynamic can be viewed as a discrete coalgebra, that is, a rule that, given the current state, produces the next one. Biologically, a coalgebra describes how a living system transforms or replicates, generating its future state from the present.

#### 9.2.1.2 Definition (Lineage Category $R_\alpha$ )

##### Level-0 Objects (Informational Matter)

$$Ob_0(R_\alpha) = \{X(0), \alpha(t), \dots\}$$

These are the concrete information handlers.  $X(0)$  is the original handler from which all others derive. Each  $\alpha(t)$  represents a discrete material state of the lineage (molecule, replicator, population, or configuration). The morphisms between these objects describe material transformations: replication, mutation, or reproduction.

##### Level-1 Objects (Functional)

$$Ob_1(R_\alpha) = \{[X(0), \alpha(1)]\} \cup \{[\alpha(t), \alpha(t+1)] | t \geq 1\}$$

For each pair  $(\alpha(t), \alpha(t+1))$ , there exists a second level object  $[\alpha(t), \alpha(t+1)]$  representing the functional space of the possible transformations from  $\alpha(t)$  to  $\alpha(t+1)$ :

$$[\alpha(t), \alpha(t+1)] \in Ob(R_\alpha).$$

Formally,  $[\alpha(t), \alpha(t+1)]$  is an internal object of the category, but biologically it represents the rules, enzymes, or functions that carry out the transformation from one material state to another.

### 9.2.1.3 Morphisms (arrows)

Each elementary morphism of lineage  $\alpha$  is denoted

$$f_{\alpha,t} : \alpha(t) \rightarrow \alpha(t+1)$$

where  $\alpha(t)$  is the information handler at step  $t$  (informational matter), and  $f_{\alpha,t}$  is the combined action of replication and mutation producing the new IH  $\alpha(t+1)$ . Here, it is not necessary to specify internal subtypes  $s$ , since the arrow refers to the entire set or distribution of the lineage at time  $t$ .

Additionally, two internal arrows are defined to link the material level with the functional one:

$$\rho_t : \alpha(t) \rightarrow [\alpha(t), \alpha(t+1)], \quad ev_t : [\alpha(t), \alpha(t+1)] \rightarrow \alpha(t+1).$$

The arrow  $\rho_t$  assigns to each resulting state  $\alpha(t+1)$  the mechanism that can reproduce it in the future; it is an abstraction of the production of enzymes or functions enabling replication. Its composition with  $ev_t$  generates the effective transformation of the system:

$$f_{\alpha t} = ev_t \circ \rho_t$$

Altogether, the system produces the means that ensure its own future replication, thus fulfilling the condition of discrete (M,R) closure:

$$(M, R) : \alpha(t) \xrightarrow{\rho_t} [\alpha(t), \alpha(t+1)] \xrightarrow{ev_t} \alpha(t+1)$$

See the Reproduction section 9.2.6.2 for more details on the replication.

### 9.2.2 Supercategory R (Category of Categories)

The set of all lineages forms a supercategory  $R$ , whose objects are the categories  $R_\alpha$  and whose morphisms are functors between them. These morphisms represent processes of reproduction or interaction between lineages (e.g., hybridization, gene transfer, symbiosis, etc.). The natural transformations between these functors, representing coherent changes, for instance, in modes or preferences of pairing, endow  $R$  with the structure of a bicategory, in which the composition of

processes is preserved not strictly but up to natural equivalence, that is, up to a structural correspondence that maintains functional coherence among different modes of interaction.

In other words, the composition of processes is preserved in such a way that the biological and evolutionary coherence of the system is maintained.

$$Ob(R) = \{R_\alpha \mid \alpha \in \Lambda\}$$

$$Mor(R) = \{F: R_\alpha \rightarrow R_\beta \mid F \text{ functor}\}$$

### 9.2.3 Populations: Deterministic Functor to Set

Each object at state  $s$  in lineage  $\alpha$  is connected to the set of copies it contains. To do this, for each lineage  $\alpha$ , we define a deterministic population functor:

$$P_\alpha: R_\alpha \rightarrow Set$$

which assigns to each object  $\alpha_s(t)$  a set of instances of the information handler, and to each morphism a deterministic function between sets.

#### 9.2.3.1 Objects

Each abstract object  $\alpha_s(t) \in Ob(R_\alpha)$  is realized as a set of temporal instances given by

$$P_\alpha(\alpha(s); t).$$

#### 9.2.3.2 Morphisms

For each elementary morphism  $f_{\alpha,s}$  in  $R_\alpha$ ,

$$f_{\alpha,s}: \alpha(s) \rightarrow \alpha(s+1),$$

the functor

$$P_\alpha(f_{\alpha,s}; t): P_\alpha(\alpha(s); t) \rightarrow P_\alpha(\alpha(s+1); t+1)$$

associates the population of subtype  $\alpha_s(t)$  at time  $t$  with the population of subtype  $\alpha_{s+1}(t+1)$  at time  $t+1$ . The increment from  $s$  to  $s+1$  reflects the appearance of at least one new variant during

replication of the subtype, so that the time step also incorporates the genetic innovation accumulated within the lineage.

### 9.2.3.3 Functorial Property

$$P_\alpha(id_{\alpha(s)}) = id_{P_\alpha(\alpha(s); t)}$$

$$P_\alpha(f_{\alpha,s+1} \circ f_{\alpha,s}) = P_\alpha(f_{\alpha,s+1}) \circ P_\alpha(f_{\alpha,s})$$

### 9.2.3.4 Cardinalities and Frequencies per Lineage

Let

$$N_{\alpha,s}(t) = |P_\alpha(\alpha(s); t)|$$

denote the number of IHs of subtype  $s$  that are still present at time  $t$ ; where

$$P_\alpha(f_{\alpha,s}; t)$$

denote the replication/mutation function.

Then define

$$N_\alpha^{tot}(t) = \sum_{s \leq t} N_{\alpha,s}(t),$$

$$p_{\alpha,s}(t) = \frac{N_{\alpha,s}(t)}{N_\alpha^{tot}(t)},$$

$$\sum_{s \leq t} p_{\alpha,s}(t) = 1$$

where

$p_{\alpha,s}(t)$  denote the current frequency of that subtype.

### 9.2.3.5 Functional Summary

The functor  $P_\alpha$  links the categorical structure of the lineage

$$R_\alpha = \{\alpha(0), \alpha(1), \dots, \alpha(T)\}$$

with its concrete realization in population sets

$$P_\alpha(\alpha(s); t) \subseteq Set$$

allowing one to define

$$N_{\alpha,s}(t+1) = r_{\alpha,s}(t)N_{\alpha,s}(t)$$

where  $r_{\alpha,s}(t)$  is the effective growth or persistence rate of each type  $s$ , including the net effect of entries and exits due to mutation or extinction of that type.

### 9.2.4 Intra-Lineage Replicator Dynamics

Given the set of substates  $\{\alpha(0), \alpha(1), \dots, \alpha(t)\}$  of lineage  $\alpha$  at step  $t$  and letting  $p_{\alpha s}$  denote the current frequency of subtype  $s$  at that time, the frequency in the next generation, considering only differences in the replication rate (i.e., with no mutation or other transmission effects), is given by

$$p_{\alpha,s}(t+1) = \frac{p_{\alpha,s}(t) \cdot r_{\alpha,s}(t)}{\bar{r}_{\alpha}(t)} \quad (\text{A1})$$

where

$$\bar{r}_{\alpha}(t) = \sum_s r_{\alpha,s}(t) \cdot p_{\alpha,s}(t)$$

is the mean replication rate in the population.

#### 9.2.4.1 Consistency between Counts and the Replicator Equation

Assuming that, in the transition from  $t$  to  $t+1$ , new objects  $\alpha_{t+1}(t+1)$  are generated, belonging to a new class  $s = t+1$ , and that some classes  $s \leq t$  may disappear (i.e., become the empty set), then

$$N_{\alpha,s}(t+1) = \sum_i N_{\alpha,i}(t) \cdot r_{\alpha,i}(t) \cdot Q_{i \rightarrow s}(t) \quad (\text{A2})$$

where  $Q$  is a stochastic transition matrix that captures the probabilities of transition between states.

Thus we obtain

$$N_{\alpha}^{tot}(t+1) = \sum_s N_{\alpha,s}(t+1) = \sum_s \sum_i N_{\alpha,i}(t) \cdot r_{\alpha,i}(t) \cdot Q_{i \rightarrow s}(t) = \sum_i N_{\alpha,i}(t) \cdot r_{\alpha,i}(t) \cdot \sum_s Q_{i \rightarrow s}(t)$$

so

$$N_{\alpha}^{tot}(t+1) = \sum_i N_{\alpha,i}(t) \cdot r_{\alpha,i}(t) = N_{\alpha}^{tot}(t) \cdot \bar{r}_{\alpha}(t) \text{ because } \sum_s Q_{i \rightarrow s}(t) = 1 \forall i$$

and dividing both sides of (A2) by  $N_{\alpha}^{tot}(t+1)$  and substituting  $N_{\alpha}^{tot}(t+1) = \bar{r}_{\alpha}(t) \cdot N_{\alpha}^{tot}(t)$  into the right-hand side, we obtain

$$p_{\alpha,s}(t+1) = \frac{\sum_i p_{\alpha,i}(t) \cdot r_{\alpha,i}(t) \cdot Q_{i \rightarrow s}(t)}{\bar{r}_{\alpha}(t)}$$

To recover the replicator equation (A1), we define an effective replication rate that incorporates the effect of mutation:

$$r_{\alpha,s}^{(eff)}(t) = \frac{N_{\alpha,s}(t+1)}{N_{\alpha,s}(t)} = r_{\alpha,s}(t)Q_{s \rightarrow s}(t) + \sum_{i \neq s} \frac{N_{\alpha,i}(t)}{N_{\alpha,s}(t)} r_{\alpha,i}(t)Q_{i \rightarrow s}(t),$$

passing to frequencies,

$$r_{\alpha,s}^{eff}(t) = r_{\alpha,s}(t)Q_{s \rightarrow s}(t) + \sum_{i \neq s} r_{\alpha,i}(t)[Q_{i \rightarrow s}(t)/p_{\alpha,s}(t)]p_{\alpha,i}(t)$$

The term  $r_{\alpha,s}^{eff}(t)$  measures the effective fitness of class  $s$ .

Thus the replicator equation (A1) now becomes

$$p_{\alpha,s}(t+1) = \frac{p_{\alpha,s}(t) \cdot r_{\alpha,s}^{eff}(t)}{\bar{r}_{\alpha}^{eff}(t)} \quad (\text{A3})$$

where

$$\bar{r}_{\alpha}^{eff}(t) = \sum_u p_{\alpha,u}(t) \cdot r_{\alpha,u}(t)$$

In this way, the effective replication rate measures the total change factor of the population, including both the descendants that remain in the same class and those that arrive from other classes.

### 9.2.5 Functional or Information Functor with Two Channels $J_r$ and $J_{rt}$

The information functor  $\text{Info}_{\alpha}$  acts on the replication morphisms  $f_{\alpha,t}$ , but its value is computed as the Jeffreys divergence between the frequency distributions induced by these morphisms through the population functo  $P_{\alpha}$ . The functor can be decomposed into two channels: the selective channel  $\text{Info}_{\alpha}^{(r)}$ , which measures the information due exclusively to selection, and the total channel  $\text{Info}_{\alpha}^{(rt)}$ , which also includes the effects of transmission.

#### 9.2.5.1 Basic notation

$$p(t) = p_{\alpha}(t) = (p_{\alpha,0}(t), p_{\alpha,1}(t), \dots)$$

$$\begin{aligned}\bar{r}_\alpha(t) &= \sum_s p_{\alpha,s}(t) \cdot r_{\alpha,s}(t) \\ r'_{\alpha,s}(t) &= r_{\alpha,s}(t) / \bar{r}_\alpha(t) \\ r'(t) &= (r'_{\alpha,0}(t), r'_{\alpha,1}(t), \dots)\end{aligned}$$

### 9.2.5.2 Update by Selection

For each state  $s$  at time  $t$  ( $s \leq t$ ), the state at  $t+1$  is updated according to (A1), here written as

$$(p_r)_s(t+1) = r'_{\alpha,s}(t) p_{\alpha,s}(t)$$

### 9.2.5.3 Update by Selection + Transmision

Here, transmission includes the effects of mutation, migration, or class flow, absorbed into an effective replication rate. For each state  $s$  at time  $t$  ( $s \leq t$ ), the state at  $t+1$  is updated according to (A3).

### 9.2.5.4 Informational Functionals

#### (a) Purely Selective Information

$$Info_\alpha^{(r)}(f_{\alpha,t}) = J(p_r(t+1), p(t)) = \sum_s ((p_r)_s(t+1) - p_{\alpha,s}(t)) \ln \frac{(p_r)_s(t+1)}{p_{\alpha,s}(t)}$$

#### (b) Total Information (Selection + Transmission)

$$Info_\alpha^{(rt)}(f_{\alpha,t}) = J(p(t+1), p(t)) = \sum_s (p_{\alpha,s}(t+1) - p_{\alpha,s}(t)) \ln \frac{p_{\alpha,s}(t+1)}{p_{\alpha,s}(t)}$$

where  $J \geq 0$  and  $J = 0$  iff the distributions at  $t$  and  $t+1$  are identical.

#### (c) Transmissional (Non-Selective) Component

For each step  $t$ , we obtain the partition:

$$Info_\alpha^{(rt)} = Info_\alpha^{(r)} + T_{trans}((\vec{p}), (\vec{r}), Q),$$

where

$$T_{trans}(f, r, Q) = \sum_s f_{\alpha,s}(t) \left[ \left( \frac{r_{\alpha,s}^{eff}(t)}{\bar{r}_\alpha(t)} - 1 \right) \ln \frac{r_{\alpha,s}^{eff}(t)}{\bar{r}_\alpha(t)} - \left( \frac{r_{\alpha,s}(t)}{\bar{r}_\alpha(t)} - 1 \right) \ln \frac{r_{\alpha,s}(t)}{\bar{r}_\alpha(t)} \right]$$

is the component due to transition effects, as a function of frequencies, replication rates, and the transition matrix  $Q$ .

If  $Q = I$  (identity) then  $T_{trans} = 0$  and  $J_{rt} = J_r$  the causes of change are then purely selective.

#### 9.2.5.5 Handling of Zeros

In the case of appearance of a new class  $p_{\alpha,s}(t)=0$ ,  $p_{\alpha,s}(t+1)>0$  (e.g.  $s = t+1$ ), the corresponding term in

$J_{rt}$

$$(p_{\alpha,s}(t+1)-0)\ln p_{\alpha,s}(t+1)/0$$

diverges. Conceptually, this can be interpreted as  $J$  jumping to infinity when innovation (new information) arises.

In the case of disappearance of a class,  $p_{\alpha,s}(t)>0$ ,  $p_{\alpha,s}(t+1) = 0$ , the object transitions to the empty set, conceptually not problematic, as it represents the loss of an informational channel.

In both cases, the problem arises only when computing  $J$  explicitly. To handle this, we apply a regularization, assuming a smoothed version of the distributions such that the support is complete (no zeros). For both distributions, we assume

$$p^\varepsilon = (1 - \varepsilon)p + \varepsilon/K \quad (1 >> \varepsilon > 0),$$

where  $K$  is the number of classes.

Thus, for any  $J$ , we can write

$$J_\varepsilon(p(t+1), p(t)) = \sum_s (p_s^\varepsilon(t+1) - p_s^\varepsilon(t)) \ln \frac{p_s^\varepsilon(t+1)}{p_s^\varepsilon(t)}$$

#### 9.2.5.6 Local Fold in $R_\alpha$

$$\text{Info}_\alpha^{(rt)}(f_{\alpha,t}) = J(p_\alpha(t+1), p_\alpha(t)),$$

$$\text{Info}_\alpha^{(rt)}(T) = \sum_{t=0}^{T-1} \text{Info}_\alpha^{(rt)}(f_{\alpha,t})$$

measures the total information generated by selection and transmission within lineage  $\alpha$ .

## 9.2.6 Reproduction $M$ , Mating Fitness $m$ and $J_{\text{repr}}$

### 9.2.6.1 Reproductive Operator (Categorical Product)

Reproduction is modeled as a binary operation (a reproductive bifunctor) between lineage objects

$$M: \alpha_s(t) \times \beta_{s'}(t) \rightarrow (\alpha\beta)(t+1)$$

where  $M$  maps pairs of parental lineages to the lineage (or set of offspring lineages) they produce.

### 9.2.6.2 Pure Replication

In the case of pure replication, reproduction occurs within a single lineage and involves no pairing or combination between different entities. Two complementary types of morphisms can be distinguished:

#### (a) Identity morphism (structural invariance)

$$id_{\alpha(t)}: \alpha(t) \rightarrow \alpha(t)$$

This morphism is structural, not temporal. It represents the self-preservation or instantaneous maintenance of the informational handler in its current state, without invoking any passage of time. Biologically, it expresses the potential of the system to remain stable, preserving its organization and information content, within a given temporal frame.

#### (b) Replication morphism (causal process)

$$f_{\alpha,t}: \alpha_s(t) \rightarrow \alpha_{s'}(t+1)$$

This morphism represents the discrete temporal transition from step  $t$  to  $t+1$ . It may or may not involve a change in the internal state  $s$  depending whether replication occurs without mutation ( $s' = s$ ), preserving the internal configuration or whether replication occurs with mutation or variation ( $s' \neq s$ ), producing a new subtype within the lineage. The morphism  $f_{\alpha,t}$  therefore captures the causal process by which the system generates its next configuration, either identical or modified, maintaining lineage continuity while potentially introducing informational change.

### 9.2.6.3 Intra-Lineage Reproduction

Within a lineage, reproduction

$$M_{\alpha;s,s'}: \alpha_s(t) \times \alpha_{s'}(t) \rightarrow \alpha_{(s \boxplus s')}(t+1)$$

is represented by the discrete combination rule  $\boxplus$ , that is, a function

$$\gamma: S \times S \rightarrow S$$

that determines the type of offspring:

$$\alpha_s(t) \times \alpha_{s+1}(t) \rightarrow \alpha_{\gamma(s, s')}(t+1).$$

For example,

$$\alpha_s(t) \times \alpha_{s+1}(t) \rightarrow \alpha_{s+1}(t+1).$$

which represents that the lineage state  $s+1$  dominates, or

$$\alpha_s(t) \times \alpha_{s'}(t) \rightarrow \alpha_{s''}(t+1)$$

represents the generation of a new subtype within lineage  $\alpha$ .

#### 9.2.6.4 Inter-Lineage Reproduction

Between distinct lineages  $\alpha$  and  $\beta$ , reproduction follows a cross-lineage generation rule

$$M_{\alpha, \beta}: \alpha_s(t) \times \beta_{s'}(t) \rightarrow (\alpha\beta)_{\gamma(s, s')}(t+1)$$

where  $\gamma(s, s')$  generates combined states between lineages.

#### 9.2.6.5 Mating frequencies

Let  $p_i(t)$  and  $p_j(t)$  denote the frequencies of object  $i$  in lineage  $\alpha$  and object  $j$  in lineage  $\beta$  at time  $t$ , respectively. The probability of random mating between them is:

$$q_{ij}(t) = p_i(t) p_j(t)$$

and the probability of mating influenced by mating fitness  $m_{ij}$  is:

$$q'_{ij}(t) = q_{ij}(t) m_{ij}(t) / M(t)$$

where

$$M(t) = \sum_{i,j} q_{ij}(t) m_{ij}(t)$$

is the mean mating fitness.

After mating, the replication rates depend on the corresponding reproductive rule:

$$\phi_{\alpha, \beta}: r_\alpha \times r_\beta \rightarrow r_{\alpha\beta}^{post} \circ \hat{m}_{\alpha, \beta}$$

In general form:

$$r_{ij}^{post}(t) = G(r_{\alpha,i}(t), r_{\beta,j}(t))$$

and, in specific cases:

$$G_L(x, y) = x, G_R(x, y) = y, G_{mean}(x, y) = (x+y)/2,$$

$$G_{geom}(x, y) = \sqrt{xy}, G_\lambda(x, y) = \lambda x + (1-\lambda),$$

$$G_{hybrid}(x, y) = I_{xy}(x+y)/2$$

where  $I_{xy} \in [0, +\infty)$  is a sexual isolation factor.

#### 9.2.6.6 Reproductive Fold (Jeffreys Divergence in Pair Space)

The deviation from random mating is measured by the Jeffreys divergence in the space of pairs:

$$J_{repr}(q'(t), q(t)) = \sum_{i,j} (q'_{ij}(t) - q_{ij}(t)) \ln \frac{q'_{ij}(t)}{q_{ij}(t)}$$

This measure can, in turn, be separated into additive components of male and female sexual selection and an assortative mating component (Carvajal-Rodríguez 2018).

#### 9.2.6.7 Decomposition by Lineage Relationship

In this formulation, reproduction is decomposed into three informational channels, corresponding to the total reproductive channel  $J_{repr}$ , the intra-lineage channel  $J_{repr,intra}$ , and the inter-lineage channel  $J_{repr,inter}$ .

##### **Intra-lineage (within $\alpha$ ):**

$$J_{repr}^{intra,\alpha} = \sum_{i,j \in \alpha} (q'_{ij} - q_{ij}) \ln \frac{q'_{ij}}{q_{ij}}$$

##### **Inter-lineage (between $\alpha$ and $\beta$ ):**

$$J_{repr}^{inter,\alpha,\beta} = \sum_{i \in \alpha, j \in \beta} (q'_{ij} - q_{ij}) \ln \frac{q'_{ij}}{q_{ij}}$$

where the indices  $i, j$  range over the coexisting substates at step  $t$ .

In this partition of reproductive information, we distinguish the total, intra-lineage, and inter-lineage channels, each defined over its own normalized distribution. Since the normalizations differ, the corresponding divergences are not additive, but each preserves probability consistency and an informational interpretation within its respective domain.

### 9.2.7 Higher Level: Global Fold in R

The global fold integrates the total information generated by replication, transmission, and reproduction.

#### 9.2.7.1 Total information at t

$$Info_{Total}(t) = J_{repr}(q'(t), q(t)) + \sum_{\alpha \in \Lambda} J_{rt}(f_{\alpha}(t+1), f_{\alpha}(t))$$

#### 9.2.7.2 Global Information over T: Purely Selective Option

$$Info_{Global}^{(r)}(T) = \sum_{\alpha, t} J_r(f_{\alpha}(t+1), f_{\alpha}(t))$$

This expression measures the accumulated information due solely to selection across all lineages and discrete time steps.

#### 9.2.7.3 Global Information over T: Total Option (Selection + Transmission)

$$Info_{Global}^{(rt)}(T) = \sum_{\alpha, t} J_{rt}(f_{\alpha}(t+1), f_{\alpha}(t))$$

This includes selection plus the transmission effects (mutation, recombination, migration) accumulated through time and across all lineages.

#### 9.2.7.4 Explicit Decomposition of the Transmission Channel

$$Info_{Global}^{rt} = Info_{Global}^r + \sum_{\alpha, t} T_{Trans, \alpha}(t)$$

$$\Delta Info_{trans}^{Global}(T) = Info_{Global}^{(rt)}(T) - Info_{Global}^{(r)}(T) = \sum_{\alpha \in \Lambda} \Delta Info_{Trans}(\alpha, T)$$

represents the total information associated with non-selective transmission effects.

#### 9.2.7.5 Total Global Information in $T$

$$Info_{Global}^{Total}(T) = \sum_{t=0}^{T-1} Info_{Total}(t)$$

which captures the complete informational content accumulated by the system due to selection, transmission, and reproduction over the interval  $[0, T]$ .

**Funding:** This was funded by Xunta de Galicia (ED431C 2024/22), Ministerio de Ciencia e Innovación (PID2022-137935NB-I00) and Centro singular de investigación de Galicia accreditation 2024-2027 (ED431G 2023/07) and “ERDF A way of making Europe”.

**Data Availability Statement:** No data were used in this study.

**Conflicts of interest:** The authors declare no conflicts of interest.

## REFERENCES

- Adami C. 2004. Information theory in molecular biology. *Physics of Life Reviews* 1:3-22.
- Adami C. 2012. The use of information theory in evolutionary biology. *Annals of the New York Academy of Sciences* 1256:49-65.
- Adami C. 2024. The Evolution of Biological Information: How Evolution Creates Complexity, from Viruses to Brains. Princeton
- Anta J, Sánchez-Dorado J. 2025. The pursuitworthiness of informational biology. *Biol Philos* 40:21.
- Archibald JM. 2024. Symbiotic revolutions at the interface of genomics and microbiology. *PLOS Biology* 22:e3002581.
- Avery JS. 2021. Information Theory and Evolution. 3rd ed. World Scientific Available from: <https://www.worldscientific.com/worldscibooks/10.1142/12668>
- Ay N, Jost J, Lê HV, Schwachhöfer L. 2017. Information Geometry. Cham: Springer International Publishing Available from: <http://link.springer.com/10.1007/978-3-319-56478-4>
- Azpeitia E, Balanzario EP, Wagner A. 2020. Signaling pathways have an inherent need for noise to acquire information. *BMC Bioinformatics* 21:462.
- Baalen M van. 2013. Biological information: why we need a good measure and the challenges ahead. *Interface Focus [Internet]*. Available from: <https://royalsocietypublishing.org/doi/10.1098/rsfs.2013.0030>
- Baez JC, Otter N. 2017. Operads and phylogenetic trees. *Theory and Applications of Categories* 32:1397–1453.
- Barbieri M. 2019. A general model on the origin of biological codes. *Biosystems* 181:11-19.
- Batstone RT. 2022. Genomes within genomes: nested symbiosis and its implications for plant evolution. *New Phytologist* 234:28–34.
- Beiler KJ, Simard SW, Lemay V, Durall DM. 2012. Vertical partitioning between sister species of Rhizopogon fungi on mesic and xeric sites in an interior Douglas-fir forest. *Mol Ecol* 21:6163–6174.
- Benítez-Benítez C, Mohan AV, Sánchez-Villegas R, Gómez-Ramos I, Valdés-Florido A, Lucek K, Slovák M, Kolář F, Leitch IJ, Luceño M, et al. 2025. Bridging micro and macroevolution: insights from chromosomal

dynamics in plants. *Front. Plant Sci.* [Internet] 16. Available from: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2025.1606450/full>

Bernhardt HS. 2012. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)a. *Biol Direct* 7:23.

Bock DG, Cai Z, Elphinstone C, González-Segovia E, Hirabayashi K, Huang K, Keais GL, Kim A, Owens GL, Rieseberg LH. 2023. Genomics of plant speciation. *Plant Communications* 4:100599.

Bonduriansky R, Day T. 2018. Extended Heredity. Princeton University Press Available from: <http://www.jstor.org/stable/j.ctvc77mbg>

Bosch TCG, Wigley M, Colomina B, Bohannan B, Meggers F, Amato KR, Azad MB, Blaser MJ, Brown K, Dominguez-Bello MG, et al. 2024. The potential importance of the built-environment microbiome and its impact on human health. *Proceedings of the National Academy of Sciences* 121:e2313971121.

Brenner S. 2010. Sequences and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences* [Internet]. Available from: <https://royalsocietypublishing.org/doi/10.1098/rstb.2009.0221>

Bürger R. 2000. The Mathematical theory of selection, recombination, and mutation. Chichester: John Wiley

Cartwright JHE, Čejková J, Fimmel E, Giannerini S, Gonzalez DL, Goracci G, Grácio C, Houwing-Duistermaat J, Matić D, Mišić N, et al. 2024. Information, Coding, and Biological Function: The Dynamics of Life. *Artificial Life* 30:16–27.

Carvajal-Rodríguez A. 2018. Non-random mating and information theory. *Theoretical Population Biology* 120:103–113.

Carvajal-Rodríguez A. 2019. A generalization of the informational view of non-random mating: Models with variable population frequencies. *Theoretical Population Biology* 125:67–74.

Carvajal-Rodríguez A. 2020. Multi-model inference of non-random mating from an information theoretic approach. *Theoretical Population Biology* 131:38–53.

Carvajal-Rodríguez A. 2024a. Unifying quantification methods for sexual selection and assortative mating using information theory. *Theoretical Population Biology* 158:206–215.

Carvajal-Rodríguez A. 2024b. On Non-Random Mating, Adaptive Evolution, and Information Theory. *Biology* 13:970.

Cheng E. 2022. The Joy of Abstraction: An Exploration of Math, Category Theory, and Life. Cambridge: Cambridge University Press Available from: <https://www.cambridge.org/core/books/joy-of-abstraction/00D9AFD3046A406CB85D1AFF5450E657>

Davies P. 2019. The Demon in the Machine: How Hidden Webs of Information Are Solving the Mystery of Life. Chicago, IL: University of Chicago Press Available from: <https://press.uchicago.edu/ucp/books/book/chicago/D/bo45084244.html>

Davies P CW and Walker, Sara Imari. 2016. The hidden simplicity of biology. *Reports on Progress in Physics* 79:102601.

Dawkins R. 1978. Replicator Selection and the Extended Phenotype. *Zeitschrift für Tierpsychologie* 47:61–76.

Dawkins R. 1999. Richard Dawkins The Extended Phenotype The Long Reach Of The Gene. Oxford University Press Available from:  
<http://archive.org/details/RichardDawkinsTheExtendedPhenotypeTheLongReachOfTheGene1999OxfordUniversityPress>

Dawkins R. 2016. The selfish gene. Oxford university press

DiFrisco J, Gawne R. 2025. Biological agency: a concept without a research program. *J Evol Biol* 38:143–156.

Dobzhansky TD. 1982. Genetics and the Origin of Species: Columbia Classics edition. Columbia University Press

Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Pirajno F, O'Neil J, Little CTS. 2017. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature* 543:60–64.

Edelaar P, Bolnick DI. 2019. Appreciating the Multiple Processes Increasing Individual or Population Fitness. *Trends in Ecology & Evolution* 34:435–446.

Fisher RA. 1930. The Genetical Theory of Natural Selection. Oxford: Oxford University Press

Frank SA. 2012. Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *J Evol Biol* 25:2377–2396.

Frank SA. 2024. Two principles of success. *Proceedings of the National Academy of Sciences* 121:e2417410121.

Frank SA. 2025. Natural selection at multiple scales. *Evolution* 79:1166–1184.

Futuyma D. 2019. Russell Bonduriansky and Troy Day. Extended Heredity: A New Understanding of Inheritance and Evolution. *Evolutionary Studies in Imaginative Culture* 3:115–118.

Garcia M, Theunissen F, Sèbe F, Clavel J, Ravignani A, Marin-Cudraz T, Fuchs J, Mathevon N. 2020. Evolution of communication signals and information during species radiation. *Nat Commun* 11:4970.

García-Sancho M. 2011. From metaphor to practices: The introduction of “information engineers” into the first DNA sequence database. *Hist Philos Life Sci* 33:71–104.

Gilbert W. 1986. Origin of life: The RNA world. *Nature* 319:618–618.

Gómez-Márquez J. 2021. What is life? *Mol Biol Rep* 48:6223–6230.

Griffiths PE. 2001. Genetic Information: A Metaphor In Search of a Theory. *Philosophy of Science* 68:394–412.

Griffiths PE, Gray RD. 1997. Replicator II – Judgement Day. *Biology & Philosophy* 12:471–492.

Haldane JBS. 1932. The Causes of Evolution. London: Longmans, Green & Co.

Hancock ZB, Lehmberg ES, Bradburd GS. 2021. Neo-darwinism still haunts evolutionary theory: A modern perspective on Charlesworth, Lande, and Slatkin (1982). *Evol* 75:1244–1255.

Hansen TF. 2018. Fitness in Evolutionary Biology. Available from:  
<https://www.preprints.org/manuscript/201804.0271/v1>

- Hardin G. 1968. The Tragedy of the Commons. *Science* 162:1243–1248.
- Hledík M, Barton N, Tkačik G. 2022. Accumulation and maintenance of information in evolution. *Proceedings of the National Academy of Sciences* 119:e2123152119.
- Hoffmeyer J. 2008. Biosemiotics: An Examination into the Signs of Life and the Life of Signs. (Favareau D, editor.). Scranton, Pa.
- Hull DL. 1980. Individuality and Selection. *Annual Review of Ecology, Evolution, and Systematics* 11:311–332.
- Jablonka E, Lamb MJ. 2006. Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life. Cambridge, MA, USA: MIT Press
- Johnson NA. 2022. Speciation: genomic sequence data and the biogeography of speciation. *Natl Sci Rev* [Internet] 9. Available from: <https://dx.doi.org/10.1093/nsr/nwac294>
- Kimura M. 1961. Natural selection as the process of accumulating genetic information in adaptive evolution. *Genetics Research* 2:127–140.
- Konopiński MK. 2020. Shannon diversity index: a call to replace the original Shannon's formula with unbiased estimator in the population genetics studies. *PeerJ* 8:e9391.
- Koonin EV. 2016. The meaning of biological information. *Philos Trans A Math Phys Eng Sci* [Internet] 374. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4760125/>
- Kullback S. 1997. Information Theory and Statistics. New edition. Mineola, N.Y: Dover Publications
- Lane PA. 2024. Robert Rosen's Relational Biology Theory and His Emphasis on Non-Algorithmic Approaches to Living Systems. *Mathematics* 12:3529.
- Leinster T. 2025. Basic Category Theory. Available from: <http://arxiv.org/abs/1612.09375>
- Letelier JC, Marín G, Mpodozis J. 2003. Autopoietic and (M,R) systems. *Journal of Theoretical Biology* 222:261–272.
- Lewontin RC. 1970. The Units of Selection. *Annual Review of Ecology and Systematics* 1:1–18.
- Lewontin RC. 1974. The genetic basis of evolutionary change. New York: Columbia University Press
- Lian C, Narimatsu M, Nara K, Hogetsu T. 2006. Tricholoma matsutake in a natural *Pinus densiflora* forest: correspondence between above- and below-ground genets, association with multiple host trees and alteration of existing ectomycorrhizal communities. *New Phytol* 171:825–836.
- Lovelock JE. 1972. Gaia as seen through the atmosphere. *Atmospheric Environment (1967)* 6:579–580.
- Lovelock JE, Margulis L. 1974. Atmospheric homeostasis by and for the biosphere: the gaia hypothesis | Tellus A: Dynamic Meteorology and Oceanography. *Tellus A* 26:2–10.
- Mac Lane S. 1971. Categories for the working mathematician. Heidelberg: Springer
- Mahadevan S. 2024. Universal Imitation Games. Available from: <http://arxiv.org/abs/2405.01540>

Margalef, R. 1968. Perspectives in ecological theory. University of Chicago Press Available from: <http://archive.org/details/Margalef>

Margalef R. 1996. Information and uncertainty in living systems, a view from ecology. *Biosystems* 38:141–146.

Margulis L. 1976. Genetic and evolutionary consequences of symbiosis. *Experimental Parasitology* 39:277–349.

Margulis L. 1991. Symbiogenesis and symbiontism. (Conference on Symbiosis as a Source of Evolutionary Innovation (1989 : : Bellagio I, editor.). Cambridge, MA: MIT Press

Margulis L, Bermudes D. 1985. Symbiosis as a mechanism of evolution: status of cell symbiosis theory. *Symbiosis (Philadelphia, PA)* 1:101–123.

Margulis L, Sagan D. 2000. What Is Life? University of California Press

Marín C. 2024. Three types of units of selection. *Evolution* 78:579–586.

Marín C, Wade MJ. 2025. Bring back the phenotype. *New Phytologist* 246:2440–2445.

Mariscal C. 2021. Life. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy. Winter 2021. Metaphysics Research Lab, Stanford University. Available from: <https://plato.stanford.edu/archives/win2021/entries/life/>

Maturana H, Varela FJ. 2005. De máquinas y seres vivos : autopoiesis : la organización de lo vivo. 2005th ed. Buenos Aires

Mazzocchi F. 2012. Complexity and the reductionism-holism debate in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 4:413–427.

Mazzocchi F. 2025. An Investigation Into the Notion of Complex Systems. *Found Sci [Internet]*. Available from: <https://doi.org/10.1007/s10699-025-09975-2>

McKay CP. 2004. What Is Life—and How Do We Search for It in Other Worlds? *PLOS Biology* 2:e302.

Miller MB, Bassler BL. 2001. Quorum Sensing in Bacteria. *Annual Review of Microbiology* 55:165–199.

Montévil M, Mossio M. 2015. Biological organisation as closure of constraints. *J Theor Biol* 372:179–191.

Moreno Á, Peretó J. 2026. Metazoans: An Explosion of Agency. In: Moreno Á, Peretó J, editors. An Evolutionary Story of Agency: How Life Evolved to Act on its Own. Cham: Springer Nature Switzerland. p. 77–97. Available from: [https://doi.org/10.1007/978-3-032-05644-3\\_6](https://doi.org/10.1007/978-3-032-05644-3_6)

Mossio M. 2024. Introduction: Organization as a Scientific Blind Spot. In: Mossio M, editor. Organization in Biology. Cham: Springer International Publishing. p. 1–22. Available from: [https://doi.org/10.1007/978-3-031-38968-9\\_1](https://doi.org/10.1007/978-3-031-38968-9_1)

Mossio M, Montévil M, Longo G. 2016. Theoretical principles for biology: Organization. *Progress in Biophysics and Molecular Biology* 122:24–35.

Mossio M, Moreno A. 2010. Organisational closure in biological organisms. *History and Philosophy of the Life Sciences* 32:269.

Mossio M, Pontarotti G. 2019. Conserving Functions across Generations: Heredity in Light of Biological Organization. *The British Journal for the Philosophy of Science* 73:249–278.

Moya A, Peretó JG. 2011. Simbiosis: seres que evolucionan juntos. Madrid: Síntesis

Newman SA, Benítez M, Bhat R, Glimm T, Kumar KV, Nanjundiah V, Nicholson DJ, Sarkar S. 2025. Agency in the Evolutionary Transition to Multicellularity. *The Quarterly Review of Biology* 100:83–118.

Nicholson DJ. 2019. Is the cell really a machine? *Journal of Theoretical Biology* 477:108–126.

Noble D. 2011. A theory of biological relativity: no privileged level of causation. *Interface Focus* 2:55–64.

Noble D, Noble R. 2023. How Purposive Agency Became Banned from Evolutionary Biology. In: Corning PA, Kauffman SA, Noble D, Shapiro JA, Vane-Wright RI, Pross A, editors. *Evolution “On Purpose”: Teleonomy in Living Systems*. The MIT Press. p. 0. Available from: <https://doi.org/10.7551/mitpress/14642.003.0015>

Noble D, Noble R. 2025. How the Central Dogma and the Theory of Selfish Genes Misled Evolutionary and Medical Sciences in Understanding Multi-factorial Diseases. *Evol Biol* 52:138–148.

Nurse P. 2008. Life, logic and information. *Nature* 454:424–426.

Okasha S. 2008. Evolution and the Levels of Selection. Oxford : Oxford ; New York

Otsuka J. 2019. The Role of Mathematics in Evolutionary Theory. *Elements in the Philosophy of Biology* [Internet]. Available from: <https://www.cambridge.org/core/elements/role-of-mathematics-in-evolutionary-theory/873B4FDDA1F95DC93D65291D318A0D69>

Oyama S. 2000. The Ontogeny of Information: Developmental Systems and Evolution. Duke University Press Available from: <http://www.jstor.org/stable/j.ctv1220mm5>

Pardo L. 2018. Statistical Inference Based on Divergence Measures. In: 0 ed. Chapman and Hall/CRC. Available from: <https://www.taylorfrancis.com/books/9781420034813>

Paredes O, Farfán-Ugalde E, Gómez-Márquez C, Borrayo E, Mendizabal AP, Morales JA. 2024. The calculus of codes - From entropy, complexity, and information to life. *Biosystems* 236:105099.

Pattee HH. 2001. The physics of symbols: bridging the epistemic cut. *Biosystems* 60:5–21.

Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Molecular Ecology* 25:2337–2360.

Prinz R. 2023. Nothing in evolution makes sense except in the light of code biology. *Biosystems* 229:104907.

Rashevsky N. 1954. Topology and life: In search of general mathematical principles in biology and sociology. *Bulletin of Mathematical Biophysics* 16:317–348.

Richards TA, Moran NA. 2024. Symbiosis: In search of a deeper understanding. *PLOS Biology* 22:e3002595.

Rosen R. 1972. Some relational cell models: the metabolism-repair systems. In: Foundations of mathematical biology. Elsevier. p. 217–253.

Rosen R. 1991. *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life*. Columbia University Press

Roughgarden J. 2020. Holobiont Evolution: Mathematical Model with Vertical vs. Horizontal Microbiome Transmission. *Philosophy, Theory, and Practice in Biology* [Internet] 12. Available from: <http://hdl.handle.net/2027/spo.16039257.0012.002>

Ruan Q, Geng S, Yu J, Lu L, Liu Y, Chen J, Liao Q, Guo R. 2026. Microbial quorum sensing: Mechanisms, applications, and challenges. *Biotechnology Advances* 86:108733.

Sherwin WB. 2018. Entropy, or Information, Unifies Ecology and Evolution and Beyond. *Entropy* 20:727.

Shirt-Ediss B, Ferrero-Fernández A, De Martino D, Bich L, Moreno A, Ruiz-Mirazo K. 2025. Modelling the prebiotic origins of regulation and agency in evolving protocell ecologies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 380:20240287.

Skene KR. 2024. Systems theory, thermodynamics and life: Integrated thinking across ecology, organization and biological evolution. *Biosystems* 236:105123.

Smith E. 2024. Beyond fitness: The information imparted in population states by selection throughout lifecycles. *Theoretical Population Biology* 157:86–117.

Sterelny K, Smith KC, Dickison M. 1996. The extended replicator. *Biol Philos* 11:377–403.

Suárez J, Lloyd EA. 2023. Units of Selection. Cambridge: Cambridge University Press Available from: <https://www.cambridge.org/core/elements/units-of-selection/326C3444079F783C808A7E44C17BFF30>

Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF, Bosch TCG, Cryan JF, Gilbert SF, Goodnight CJ, Lloyd EA, et al. 2016. Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes. *mSystems* 1:e00028-16.

Tuyérás R. 2018. Category theory for genetics I: mutations and sequence alignments. *Theory and Applications of Categories* 33:1269–1317.

Tuyérás R. 2023. Category theory for genetics II: genotype, phenotype and haplotype. Available from: <http://arxiv.org/abs/1805.07004>

Varela FG, Maturana HR, Uribe R. 1974. Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems* 5:187–196.

Vega F. 2024. The cell as a semiotic system that realizes closure to efficient causation: The semiotic (M, R) system. *BioSystems* 240:105226.

Wagner A. 2007. From bit to it: How a complex metabolic network transforms information into living matter. *BMC Systems Biology* 1:33.

Wagner A. 2017. Information theory, evolutionary innovations and evolvability. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372:20160416.

Walker SI, Davies PCW. 2013. The algorithmic origins of life. *Journal of The Royal Society Interface* 10:20120869.

Wang R, Meng Q, Wang X, Xiao Y, Sun R, Zhang Z, Fu Y, Di Giuseppe G, Liang A. 2024. Comparative genomic analysis of symbiotic and free-living *Fluviibacter phosphoraccumulans* strains provides insights into the evolutionary origins of obligate Euplotes-bacterial endosymbioses. *Applied and Environmental Microbiology* 90:e01900-23.

Week B, Russell SL, Schulenburg H, Bohannan BJM, Bruijning M. 2025. Applying evolutionary theory to understand host-microbiome evolution. *Nat Ecol Evol*:1-12.

Westall F, Brack A, Fairén AG, Schulte MD. 2023. Setting the geological scene for the origin of life and continuing open questions about its emergence. *Front Astron Space Sci* 9:1095701.

Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97-159.

Wu Y. 2023. A Category of Genes. Available from: <http://arxiv.org/abs/2311.08546>

Yockey HP. 2005. Information theory, evolution, and the origin of life. New York: Cambridge University Press

Yurchenko SB. 2023. Is information the other face of causation in biological systems? *Biosystems* 229:104925.

Zilber-Rosenberg I, Rosenberg E. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev* 32:723-735.