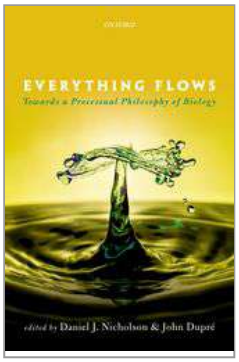


Everything Flows: Towards a Processual Philosophy of Biology
Daniel J. Nicholson and John Dupré

Print publication date: 2018
Print ISBN-13: 9780198779636
Published to Oxford Scholarship Online: July 2018
DOI: 10.1093/oso/9780198779636.001.0001

Intersecting Processes Are Necessary Explanantia for Evolutionary Biology, but Challenge Retrodiction

Eric Bapteste
Gemma Anderson

DOI:10.1093/oso/9780198779636.003.0014

Abstract and Keywords

Processes are ubiquitous in biology and play a key explanatory role in evolutionary biology, where they are frequently depicted by patterns. In particular, phylogenetic trees represent divergence from a last common ancestor with a branching pattern. However, the increasingly recognized underdetermination of phylogenetic trees limits the accuracy of tree-based retrodiction. Even phylogenetic networks, which include additional processes intersecting with vertical descent, still provide incomplete descriptions of evolutionary processes, as they usually miss processes that impact unrelated lineages. Interaction networks highlight the intersection of processes that sustain biological diversity. The complex topology of all these networks further challenges retrodiction. Remarkably, when intersecting processes are involved in evolutionary transitions, they introduce new biological processes on Earth. Processes, and hence the explanantia of evolutionary biology, evolve, which challenges uniformitarian approaches to retrodiction. Despite these difficulties, a yet to be introduced typology of processes would help to analyse the (big) processual picture of life.

Keywords: eukaryogenesis, evolution, genomics, lateral gene transfer, microbiology, network, process, tree of life, web of life

1. Introduction

Ecosystems change, species transform, humans develop and age, tomatoes rot, biofilms grow, genes mutate and recombine: processes are everywhere in biology. Responsible for both stasis and change, they affect and effect multiple levels of biological organization. Essentially, every biologist is engaged in the description of processes. The study of biological evolution is itself largely a study of processes.

Frequently, processes constitute natural explanantia in evolutionary biology (and this is probably true of other historical sciences as well). ‘Divergence from a common ancestor’ is a well-known explanans for biodiversity. Processes are often described by simple names or phrases such as ‘speciation’, ‘eukaryogenesis’, ‘endosymbiosis’, ‘descent with modification’, and so on. Yet one cannot help but be struck by the actual complexity of the causal interactions that such terms refer to, if one looks beneath the simple terminology. While metaphysical inquiries into the nature of biological processes may be out of the reach of standard scientific discourse, epistemic approaches to such processes can nevertheless be attempted. In particular, processes can be captured by patterns upon which a scientific discourse can be built.

For example, in phylogenetics, a tree-like (or branching) pattern usually captures a tree-like process of evolution—that is, a series of divergences from a last common ancestor—as changes accumulate within a lineage. Generations of evolutionary biologists have searched for the actual succession of these branching patterns in order to illustrate the split from an ancestral form into novel lineages. This has led to attempts to reconstruct the universal tree of life. Phylogenetics has unquestionably become a central practice in evolutionary biology, and the tree-like pattern a gold standard in investigations of evolutionary processes.

At the same time, however, biologists also describe living things as organized assemblies of more or less interdependent components, abstracted as intertwined and **(p.284)** interconnected regulatory, metabolic, protein–protein interaction, genetic, and developmental networks (Alon 2006; Wilkins 2007; Yafremava et al. 2013). Thus, focusing on the evolution of the relationships between organismal components offers another strategy to explain organismal evolution. This perspective, inspired by the science of evolving networks, requires treatments that are complementary to phylogenetic analyses, as the latter are not designed to model a plurality of intersecting processes operating at distinct timescales on interdependent components.

In this chapter we argue that intersecting processes are an important, underestimated explanans for the evolution of biodiversity, but that their due consideration will challenge retrodiction. In section 2 we review how phylogenetics has popularized the study of descent with modification by using a

branching (tree-like) pattern, commonly used for retrodiction. However, since the same phylogeny is often compatible with very different processes, tree-based approaches only provide a partial and ambiguous account of evolutionary history. Trees leave out essential intersecting processes that are part of the explanation of evolutionary history. Moreover, the integration of these latter processes into evolutionary studies requires other tools, methods, and patterns than those currently explored in evolutionary trees.

In section 3 we argue that even phylogenetic networks, such as Doolittle's famous web of life, are incomplete representations of the intersecting processes that affect organismal and viral lineages. Current phylogenetic network approaches still underestimate intersecting processes that affect more than one class of closely phylogenetically related agents (e.g. cells, or viruses, or plasmids). To deal with this, we introduce an expanded and updated version of Doolittle's web of life. In our version, gene externalization—the process by which a copy of a gene is placed in an unrelated genome host structure—receives a specific representation. This process implies that genes get disseminated across genetic exchange communities. It thus becomes impossible to infer past history by tracing such a complex network to a single root node. This multiplicity of connections between genomes encourages instead community-level analyses for retrodiction. However, beyond the development of new network patterns, alternative exploratory strategies are still needed to account for additional intersecting processes.

In section 4 we argue that this goal can be achieved by studying the evolution of biological organizations, abstracted as interaction networks that feature elements that can be of different types, of different origins, and from distinct levels. Crucially, the evolution of such biological organizations can be causally decoupled from speciation events. Thus, the branching pattern used to make sense of evolution will not naturally describe or explain changes in these networks, which are sustained by intersecting processes. By contrast, we briefly illustrate how network patterns have been used in studies of molecular evolution. These examples show how evolutionary studies benefit from topological explanations, as there is an important link between the phenotypes that evolutionary biologists seek to explain and the evolution of the organization of the networks that cause these phenotypes. This practice pertains to the network sciences, and more precisely to the study of evolving networks.

In section 5 we complete our epistemic diagnosis that intersecting processes are missing from the explanatory toolbox of evolutionary biology by discussing a **(p. 285)** fundamental challenge raised by these processes. Since processes evolve and since they are our explanantia, then the latter also evolve. The merging of processes, creating new forms of organization and novel processes, is a common feature of evolutionary transitions, and is therefore a major challenge for retrodiction. These challenges notwithstanding, we predict an increased use of

evolving and phylogenetic networks in evolutionary biology. As a result, we propose that the use of additional network patterns to explain the evolution of biological phenomena from the molecular to the organismal level could be achieved by developing a (yet to be introduced) network-based typology of evolutionary processes.

2. An Increasingly Appreciated Issue: The Underdetermination of Phylogenetic Trees

Phylogenetics has popularized the study of descent with modification by using a branching (tree-like) pattern (Felsenstein 2004). Tree reconstruction can be achieved for entities from different levels of biological organization, from molecules to organisms. Thus there are gene trees, protein trees, genome trees, and species trees flourishing in the literature. A common use of these phylogenetic patterns is retrodiction, that is, making 'predictions' about the past. For example, the relative order of species divergence is often inferred from gene trees, because branchings in single-copy gene trees are explained as speciation events. However, a gene tree topology is not always so easy to interpret in processual terms (i.e. as speciations; see Lapointe et al. 2010). The reason for this is that the same gene tree is often compatible with very different processes. Consequently, a tree topology provides neither a sufficient nor a complete explanation of the complexity of biological evolution. Rather than offering unequivocal evidence about the actual course of evolutionary events, gene trees provide inconclusive evidence. They require extra-phylogenetic assumptions about the likelihood of various types of evolutionary processes to be interpreted, such as, for example, the likelihood of speciation event versus the likelihood of multiple gene transfers (Ku et al. 2015a; Ku et al. 2015b).

This situation is nicely treated in the work of Ku and colleagues (Ku et al. 2015b), which is concerned with defining the correct interpretation of the observed, yet puzzling phylogenetic positions of eukaryotic sequences in gene trees, which mix sequences from prokaryotes (archaea and bacteria) with eukaryotic sequences (Figure 14.1). Since archaea, bacteria, and eukaryotes belong to three distinct domains, had their genes evolved separately within each of these lineages, it is

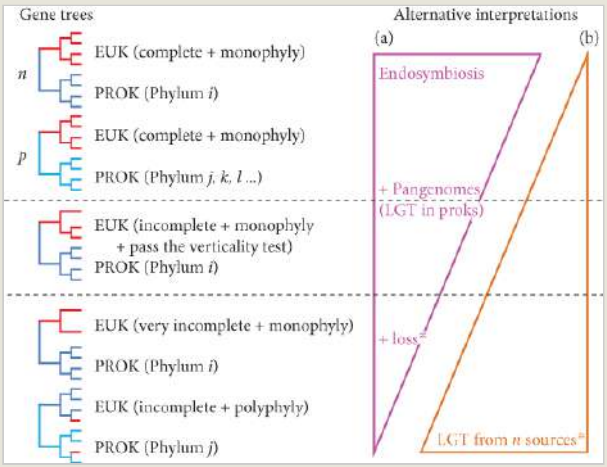


Figure 14.1 Alternative interpretations of the same gene tree, and sets of gene trees. Phylogeneticists work with gene trees represented on the left (the branches are coloured in red for eukaryotic sequences and in blue for

extremely improbable that eukaryotic sequences would be mixed with prokaryotic sequences in the gene trees. It thus becomes necessary to invoke introgressive processes (i.e. processes in which the genetic material of a particular evolutionary unit propagates into different host structures and is replicated within those host structures) or reconstruction artefacts designed to make sense of these complex topologies. Some introgressive processes must be added to speciation to explain why some eukaryotic genes look like prokaryotic genes. But that recognition raises another question: what introgressive process is the relevant explanans for these puzzling phylogenetic patterns?

prokaryotic sequences). Gene trees are organized (from top to bottom) by increasing interpretative complexity, according to the messiness and patchiness of the eukaryotic sequences and their sources. Each topology can be interpreted by appealing to different processes, indicated on the right. Interpretations are not mutually exclusive, but their likelihoods differ in accordance with the likelihood of the processes that are deemed to be the explanantia. The width of the each triangle indicates the relative likelihood of the interpretation of the gene trees on the left. The * stresses that, to favour (a), a process must explain massive independent gene losses in eukaryotes; whereas, to favour (b), lateral gene transfer (LGT) by vectors must appear likely, and hence processes of eukaryote-to-eukaryote or prokaryotes-to-eukaryotes gene transfers must be identified.

(p.286) The answer to that question is hardly ever in the gene tree. Methods with limited flexibility in the patterns they can display, such as phylogenetic tree reconstruction, cannot be used uncritically to infer these processes (Bapteste et al. 2012). As Ku et al. (2015b) put it: ‘[t]here are at least two competing alternatives to account for prokaryotic genes in eukaryotes—gradual LGT accrual versus episodic gene transfer from organelles’. One process is slow and can involve multiple gene donors (Keeling and Palmer 2008), the other is fast and involves very few donors.[1] This example highlights a fundamental reason for the failure of the branching pattern to serve as a universal explanans in evolutionary biology: trees leave out essential features of evolution in the real world, namely intersecting processes (Bapteste et al. 2009). However, the example also shows that the inclusion of such processes is not straightforward.

(p.287) 3. Intersecting Processes Are Also Absent from Phylogenetic Networks

Phylogenetic tree reconstruction methods are not the only phylogenetic approaches facing a need to better integrate intersecting processes. Phylogenetic networks, representing lateral gene transfer (LGT) and

hybridization by lateral edges between diverging branches, also stand to be improved in this respect.

The observation that intersecting processes affect more than one class of closely phylogenetically related agents (e.g. cells, or viruses, or plasmids) is still underappreciated. The famous diagram by Ford Doolittle represents LGTs and endosymbioses across the web of life by a multi-rooted network, in which the three cellular domains progressively emerge from complex ancestral populations (Doolittle 1999). This diagram is often intended to illustrate in an extreme way how much complexity intersecting processes could produce if introgressive processes were massive and widespread (Huson et al. 2010). But in many respects this heuristic drawing is not in the least bit extreme. It does not feature mobile genetic elements (MGEs), which are key players in the introgressive processes that intersect with vertical descent (Halary et al. 2010). Large quantities of gene sharing occur between MGEs and cells, between cells, and between MGEs (Bapteste et al. 2012; Bapteste 2014; Halary et al. 2010; Jachiet et al. 2014; Yutin et al. 2013). Moreover, Doolittle's drawing does not feature the relationships between hosts and microbiotas known to produce holobionts, which presumably span across the eukaryotic web of life. It does not show, more generally, that all animals, plants, fungi, and many protists are associated with prokaryotes in ways that affect their development and evolution (Brucker and Bordenstein 2012; Gilbert et al. 2012). Therefore an expanded (updated) version of Doolittle's web of life is now required. Figure 14.2 offers a first step in this direction.[2]

Figure 14.2 stresses the intersection of unrelated, evolving lineages: organisms and MGEs. It highlights the genetic intertwining between viral and cellular genomes (Bapteste and Burian 2010; Bapteste et al. 2012; Filee et al. 2006; Filee 2014; Raoult 2009; Villarreal and Witzany 2010), in addition to classic horizontal gene transfer. The red lines highlight an additional, overlooked intersecting process, which we call 'gene externalization'. During gene externalization, a copy of a gene is placed in the genome of an unrelated host. As a result, genetic material from a

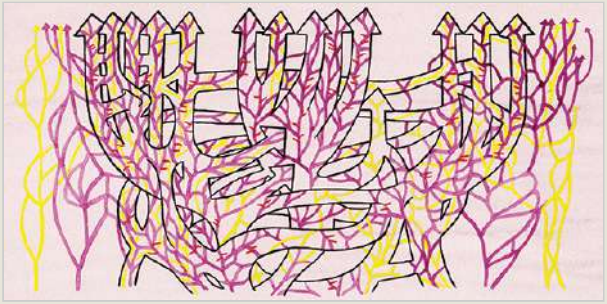


Figure 14.2 Adaptation of Doolittle's diagram (Anderson, G., watercolour on paper, 2015). The famous diagram by Ford Doolittle, slightly updated to represent transfers between archaeal and bacterial lineages, is delimited by a black outline that represents cellular lineages. It shows the vertical and lateral processes of evolution at the origin of cellular diversity. Purple and yellow lines (corresponding to the vertical and lateral

evolution of viral and plasmid lineages, respectively) complement and expand this classic drawing of the web of life. Red lines indicate gene externalization events between cellular organisms and mobile genetic elements.

particular genome or from a particular lineage disseminates, leaving copies in numerous other genomes or lineages beyond its original lineage. By this process, copies of genes feature outside their genome of origin, constituting a pool of externalized genes. Such gene externalization is not the typical LGT event found in textbooks, nor is it adequately represented in phylogenetic networks by an arrow between a cellular donor and a cellular recipient. Such an exemplar LGT requires two steps of gene externalization. A gene transfer between two cells mediated by a MGE presupposes movement from the bacterial genome to a MGE, and also from that MGE to the next bacterial genome. **(p.288)** Gene externalization can thus be seen as one half of such a schematic LGT. This difference between gene externalization and LGT means that rules (usually regarding functional biases) discovered about LGT may differ from rules about gene externalization.[3,4] Invisible in phylogenetic networks, gene externalization requires its own specific representation.

Importantly, these considerations regarding the intersection between vertical descent, LGT, and gene externalization suggest an alternative strategy for retrodiction. Rather than assuming that homologous genes coalesce (i.e. trace back to a single ancestral copy in one common ancestral genome; see Dagan and Martin 2007), they suggest an opposite, complementary perspective. Genes disseminate and genomes dissipate across genetic exchange communities (Skippington and Ragan 2011), and retrodiction thus requires community-level analyses of molecular evolution (Bapteste 2014; Corel et al. 2016). This might offer a valuable opportunity. Gene externalization through MGEs may be as old as viruses. Consequently, extant viral genomes may have preserved some early informative imprints about the history of **(p.289)** life inherited from past viral genomes (Filee et al. 2003). The focus on genome decoalescence therefore invites evolutionary biologists to dig into an unusual record of molecular evolution—unusual, yet potentially the largest such record: the copies of externalized genes present in MGEs. Overall, a comprehensive investigation of intersecting processes is likely to require tools, methods, and patterns that complement phylogenetic approaches. We will now show what alternative explanatory strategies are actually being explored in evolutionary studies.

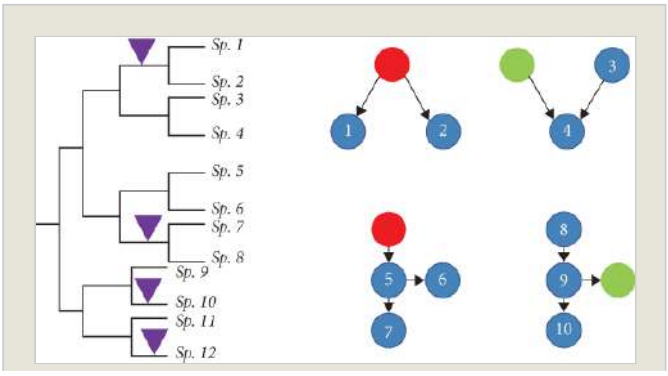
4. The Need to Investigate Reticulate Intersecting Processes in Evolutionary Studies

Intersecting processes such as merging (Bapteste et al. 2012; Méheust et al. 2015), autocatalytic cycles (Eigen 1971; Eigen and Schuster 1977, 1981), feedback loops (Milo et al. 2002), and fast-forward loops (Alon 2006) are increasingly being included in evolutionary theory. The recognition of the key

explanatory role played by these processes—captured epistemically by merging patterns, cycles, and motifs in interaction networks—is being paralleled by the realization that biological systems are organizations. Specifically, biological systems can be modelled as structured and dynamic sets of interacting components. Accordingly, molecules and organisms are increasingly considered (i) to be evolving as parts of interaction networks, and/or (ii) to be themselves composed of intersecting processes represented by intertwined networks that describe their intimate organizations (Braillard 2008).

The theoretical importance of the concept of organization in biology has long been recognized (Maturana and Varela 1980; Alon 2006; Wilkins 2007; Nicholson 2012). Crucially, the evolution of biological organizations can be causally decoupled from speciation events. More precisely, interacting organismal components can be of different ages, some having appeared earlier than others. As such, organisms are ‘temporal mosaics’, as is often illustrated by atavisms. Moreover, components may have different persistence spans—more stable components introduce structural and functional continuity in organismal evolution. Interestingly, well-adjusted biological organizations may persist over longer evolutionary periods than that of the given species from which they initially derive. For example, certain reptilian jaw-joint bones evolved into two of the middle-ear bones of different mammalian lineages. Importantly, interactions, and eventually interdependencies, between components can change over time. Preexisting components can get recycled and used to fulfil novel functions, as illustrated in co-options and evolutionary tinkering events. However, the standard epistemic branching patterns used to make sense of lineage divergences from a common ancestor will not naturally describe or explain changes in these networks. As a result, including interaction networks in comparative analyses has the potential to enhance our explanations of biodiversity. Typically, when intersecting processes are considered, one realizes that another form of selection is at play in the history of life. Selection—not to mention neutral processes—does not act only between organisms, but also within organisms (Bouchard 2014; see Figure 14.3). Relationships between components of interaction networks constrain the heritability and the variation of these components.

(p.290) Therefore, it is also compelling to study the evolution of biological organizations (composed of elements that can be of different types, of different origins, and at distinct levels) as the result of intersecting processes. We can illustrate this



move towards a more inclusive study of evolution—one that includes intersecting processes—by considering two emblematic examples at the molecular level.

4.1. Explaining the evolution of translation with a hypercycle

In the 1970s Manfred Eigen relied on intersecting processes to offer a historical explanation of the evolution of translation (Eigen and Schuster 1977; Eigen et al. 1991). He sought to understand how the instructions of nucleic acids became translatable into proteins during the early history of life.[5] Specifically, Eigen tried to solve the following conundrum: present-day biosynthesis relies upon two different types of biological entities: enzymes and nucleic acids acting in a complex network of reactions and defining a composite macromolecular system. In that regard, the origin **(p.291)** of biosynthesis is a type of chicken–egg problem, as in this system nucleic acids are causally responsible for enzymes and enzymes are causally responsible for nucleic acids. It is thus very difficult to imagine which came first, especially since this process, biosynthesis, evolved several billion years ago. Eigen used two major intersecting processes, autocatalytic loops and self-instructive cycles, to explain the evolution of nucleoproteic biosynthesis from a chaotic mix of molecules.

Figure 14.3 Two complementary explanatory frameworks. The phylogenetic framework on the left panel provides a standard form of evolutionary explanation. Here properties are mapped against the species phylogeny. The purple triangle represents a recurring trait, produced by convergence or parallelism, for example the loss of a flagellum. The distribution of this character, which is at odds with the species phylogeny, suggests multiple independent occurrences during evolution, and as such is not explained by the phylogenetic tree *sensu stricto*. The network framework on the right panel provides an alternative form of evolutionary explanation. Here the nodes represent agents, for example genes, and the arrows represent interactions between them. Red-coloured nodes represent agents which, should they be lost (for example due to the fact that many mutations can make a given gene disfunctional), have a high likelihood of affecting downstream agents owing to the structure of the network, and this can result in convergent evolution. The loss of a feature in a species can be explained by the topology of the interaction network, and its repetition, by the commonality of this topology. By contrast, green nodes could be lost without major effects on the rest of the interaction network.

Eigen introduced as explanans an interaction network organized as a ‘hypercycle’ (Figure 14.4), which integrates the different processual behaviors of both kinds of molecules: autocatalysis for enzymes, self-instructions for nucleic

acids. He insisted that his solution brought about a novel type of process, with fundamentally new evolutionary properties:

> It is the object of this paper to show, first that the breakthrough in molecular evolution must have been brought about by an integration of several self-reproducing units to a cooperative system and, second, that a mechanism capable of such an integration can be provided only by the class of hypercycles.
>
> (Eigen and Schuster 1977: 00)

In the hypercycle, as soon as self-instructive nucleic acids produce some enzyme involved in the replication of the nucleic acids, any advantageous mutation in the nucleic acids that would produce a more effective enzyme will be preserved in the system. When enzymes get improved, they become more efficient at copying nucleic acids; thus the error copy rate in nucleic acids decreases and longer nucleic acids can evolve, which enhances the information content of the system. Hypercycles can not only preserve their original information content but also enlarge it and stabilize it. **(p.292)** The patterns of cycles (autocatalytic loops and self-instructive loops) and of hypercycles were thus mobilized to construct a theory about nothing less than the evolution of translation in early life.

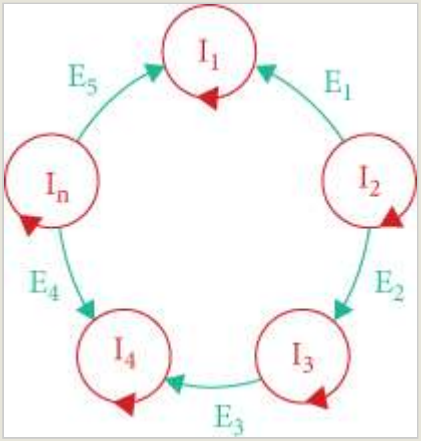


Figure 14.4 The hypercycle (Anderson, G., from Eigen 1977). A catalytic hypercycle consists of self-instructive units Ii with two-fold catalytic functions. As autocatalysts or—more generally—as catalytic cycles, the intermediates Ii are able to instruct their own reproduction and, in addition, provide catalytic support for the reproduction of the subsequent enzymatic intermediate. The figure depicts the cyclic hierarchy. Eigen illustrated the stabilizing property of the

4.2. Explaining the evolution of biological functions by network analyses

Molecular biologists are increasingly using networks such as protein–protein interaction networks and regulatory networks, as well as the patterns within them (Alon 2006; Milo et al. 2002), to make sense of the evolution of specific biological functions.[6]

They believe that investigating the evolution of a biological function requires more than the phylogenetic study of a gene or protein family, in other words, more than the mapping of substitutions on a gene or a protein tree. The reason for this belief is that most of the time the old equations ‘one gene, one protein’ and ‘one protein, one function’ simply do not hold. The way functions evolve and are selected depends on intersecting processes. Causal interactions between molecules belonging to distinct gene and protein lineages sustain functional modules. Relations of interdependency and mutual selective pressures are common (Figure 14.5).

hypercycle as a whole in his simulations. He split a sentence into four words and considered that each of the words was a distinct quasi-species. The four words were thus self-replicating with similar fitness. As these four quasi-species were in competition, only one quasi-species, hence one word, would get fixed as the simulation progresses. When these words are in functional linkage, producing a chain, all the benefit from the linkage goes to the last word, and only this last word is reproduced. (It is the famous last word!) When the words are organized in a hypercycle, the entire sentence gets stabilized. This is why hypercycles have such a remarkable potential role in evolution.

Thus, a complete understanding of the origin of a function—and, by extension, of the evolution of a function—requires knowledge about the topology of the protein–protein interaction network involved in the folding of that enzyme. When this **(p.293)** enzyme requires other ‘molecular robots’ to perform its tasks, it also requires knowledge of the protein–protein interaction network to which the folded focal enzyme contributes (Alon 2006).[7] Since an enzyme function seems to be in part determined by the relational properties of this enzyme, the evolution of a protein function must be described by retracing the evolution—the topological changes—of a network that approximates such intersecting processes.[8]

Overall, evolutionary studies benefit from topological explanations because there is an important link between the phenotypes that evolutionary biologists seek to explain and the evolution of the organization of the networks that caused these phenotypes. This line of thought pertains to network sciences and, more precisely, to the study of evolving networks.

5. Processes, and Hence Explanantia, Evolve

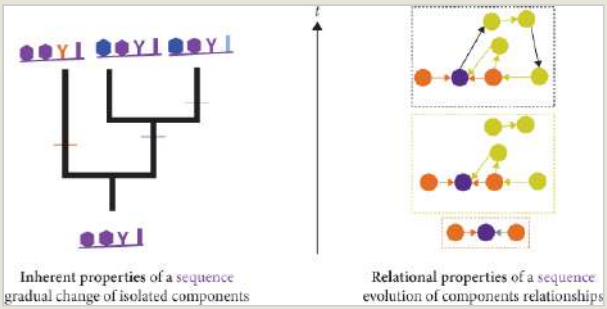


Figure 14.5 The ins and outs of phylogenetic modelling. The figure is polarized from past (bottom) to present (top). The focal enzyme is represented in purple. The left panel shows how the evolution of a function would be described in a classic framework of molecular phylogeny. The function of the purple protein evolves as substitutions accumulate in the coding genes, as represented by a protein tree. The right panel shows how the evolution of a function would be described from an evolving network perspective. Each node is a protein; an arrow represents a causal interaction between two proteins.

While our epistemic diagnosis indicates that evolutionary biology should afford greater attention to intersecting processes, these pose a fundamental challenge, which we shall consider in this section. The merging of processes, creating new organizations and novel processes, is a common feature of evolutionary transitions. Therefore, accounting for the robust merging of preexisting processes is critical for the explanation of evolution. We have already encountered this issue when discussing Eigen's theory on the origin of translation. Eigen described hypercycles as responsible for a major breakthrough in molecular evolution (Eigen and Schuster 1977). Life as we know it could not have been explained before translation evolved. However, translation was not always an active biological process on Earth. The evolution of translation was a major transition, which completely transformed biological evolution (Szathmáry 2015). What this means is that, when dramatically new processes are introduced on Earth, so are novel types of explanantia for biodiversity.[9] If Eigen is right, this was the case when autocatalytic cycles and self-replicative loops became integrated in the form of hypercycles.

Importantly, this is not the only proposed instance of a major transition coupled to the emergence of a new biological process achieved through merging (Szathmáry 2015). Although we probably do not know the exact details through which meiosis—that is, sexual reproduction—evolved, William Martin, a major evolutionary biologist, must be credited with stressing the importance of the evolution of this fundamental process during the history of life (Ku et al. 2015a). Meiosis is a process that distinguishes eukaryotes (which perform it) from prokaryotes. Meiosis is fundamental **(p.294)** because it generates genetic variability, and consequently it fuels evolution with novelties upon which natural selection can later act. As a result, the tempo and modes of evolution are

tremendously different in meiotic (sexual) and ameiotic organisms: the latter present pangenomes, while the former do not (Ku et al. 2015a). Causal explanations of biodiversity must change (i.e. they cannot rely on the same sets of processes) before and after the evolution of meiosis, because this process was not always active.

Why does it matter that meiosis probably evolved from intersecting processes? Meiosis seems to be as old as eukaryotes, and this major domain of life is almost certainly the result of the merging of (at least) two prokaryotic lineages: one ancestral archaeon and one ancestral bacterium, whose remnants can be found in the form of the mitochondrion within the cell of most present-day eukaryotes (McInerney et al. 2014; Williams et al. 2013). Thus, the process responsible for the life cycle of an early bacterium and the process responsible for the life cycle of an early archaeon merged and integrated with each other and, over time, that fusion produced a novel process, responsible for the life cycle of a novel composite organization: the eukaryotic cell. In these new life forms (from which all animals, plants, algae, amoebae, fungi, and other protists derive), meiosis contributed the necessary genetic variations in a fundamentally new way, thus preventing the extinction of eukaryotes that was expected as a result of a Muller’s ratchet (had eukaryotes evolved clonally).

Of course, this kind of merging event, however major in its evolutionary scope, is impossible to model with a tree-like pattern. It can, however, be represented with a network (see Figure 14.5). What we want to stress here is the crucial epistemic consequence of the merging of intersecting processes during eukaryogenesis. By introducing meiosis among living things, this merging scenario challenges the simple retrodiction practice that is often associated with phylogeny and its tree-like pattern. To illustrate this, imagine that there were omniscient biologists who lived at the time before meiosis emerged on Earth, and who were perfectly aware of all the biological processes of their time. The predictions of these biologists concerning future biodiversity would in all likelihood have been wrong, because a genuinely new process as fundamental as the mode of generation of variation—the invention of sex—profoundly changed the course of organismal evolution in completely unexpected and almost unimaginable ways. Obviously, these primeval biologists did not exist, but present-day biologists face a comparable challenge. In order to draw inferences about the past, they use all their understanding of processes known today. They rely on uniformitarianism, the notion that yesterday’s evolutionary processes were the same as today’s processes. That notion, however, cannot be applied uniformly across the history of life because, before eukaryotes evolved, the modes of generation of genetic variation were very different from what they are now.

New processes make it difficult to predict and retrodict life's evolution. It may even be the case that new processes add to or eliminate preexisting processes. Moreover, since the processes changed over time, the patterns used to infer evolution should also change. Using similar patterns as proxies for evolutionary processes before and after the evolution of meiosis could be very misleading, since modes and regimes of molecular variation changed profoundly around that time. The evolution of meiosis modified the very process of speciation. Thus, even those who are tempted to relate all life forms by placing them on a single tree of life should not be naïve about the processual interpretation of a so-called universal branching pattern. **(p.295)** A branch leading to some prokaryotic taxon captures very different processes from a branch leading to some eukaryotic taxon in a tree of life (Bapteste and Dupré 2013). Phylogeny is underdetermined: a diversity of processes is hidden behind a unity of patterns. Arguably, this underdermination comes at the expense of biological knowledge. When one looks at an evolutionary tree with a uniformitarian mind, one is not so strongly compelled to identify transitions; rather, one is more likely to be (overly) inclined to see what looks like a *bona fide* gradual change.

This kind of issue (i.e. conflating processes due to the use of overly simple, apparently unifying representations) may explain why other major mergings of lineages (Nelson-Sathi et al. 2012; Nelson-Sathi et al. 2015) have remained unnoticed with trees of prokaryotes (Abby et al. 2012). Mergings of lineages were recently proposed to have affected—and possibly produced—all main archaeal groups. This is especially clear in the case of Haloarchaea, whose genomes are loaded with introgressed genes of bacterial origin. These findings about archaea are particularly exciting, since they propose that the merging of metabolic genes of bacterial origin would be a common, recurring theme in the history of life, and this hints at the possibility that even intersecting processes follow some rules (Méheust et al. 2015). Even if these proposals of additional major transitions (in the case of haloarchea, from an anoxic to an oxic lifestyle) remain controversial, debates about the bacterial content of archaeal genomes are sufficient to demonstrate that intersecting processes such as LGTs, which lead to the introgression of genes into genomes, practically challenge retrodiction (López-García et al. 2015; Nelson-Sathi et al. 2012; Nelson-Sathi et al. 2015).

6. Conclusions: Towards a Typology of Processes

In this chapter we have discussed the diversity of processes with explanatory value for evolutionary biology that go beyond, and yet are complementary to, the classic notion of 'divergence from a last common ancestor'. We have argued that some of these evolutionary processes may be more appropriately described by patterns of representation that are different from the traditional branching ones. Our considerations are intended, not to belittle the importance of phylogenetic reconstructions, but to stress the need for a further integration of network concepts into evolutionary analyses (Wilkins 2007) in order to account for

intersecting processes. While the main objects of study in phylogenetics are lineages, the main objects of study in systems biology are organizations. The evolution of the networks that make up the living world is a central theme of systems biology, but it also lies at the forefront of evolutionary research, as inferring the evolution of these organizations is a complementary way to understand the evolution of life on Earth. We believe that a synthesis between phylogenetic studies and analyses of interaction networks would be highly fruitful, since evolution depends on changes in organizations as well as on the divergence and merging of lineages.

Moreover, an increased use of networks in evolutionary biology could be coupled to the development of a yet to be introduced typology of processes designed to analyse the (big) processual picture of life. In reaching this end, evolutionary studies may benefit from identifying simple patterns in evolving networks and in phylogenetic networks. This enterprise could help with translating the principles of systems **(p.296)** biology and network theory into an abstract, unifying language for theoretical biology in order to improve our understanding of the (big) processual picture of life. Typically, when systems biology seeks to identify common guiding principles in interaction networks that represent all sorts of processes—for example from transcriptional regulation to protein-protein interaction, or cellular communication (Alon 2006)—it tries to decompose large graphs into smaller meaningful motifs and modules. Using networks as a new level of abstraction to describe biological reality beyond a list of individual entities, systems biology searches for new regularities in biology (Braillard 2008). These regularities, found in the network topology, are expected to provide a universal ‘alphabet’ of interaction networks and to reveal a ‘periodic table’ for functional regulatory circuits (Kitano 2002). These small motifs, the components of networks, are detectable and are already proving to be useful in scientific inquiry. Of course, such simple patterns are abstractions, in the sense that they can be understood as smaller motifs composing a larger network of interconnected processes.[10] Yet graphlets of interaction nonetheless constitute an underappreciated regularity in biological systems.

We wish to propose the same sort of enterprise in order to identify some regularities at an even higher level of abstraction: the networks of processes, using a process typology. If successful, the most important payoff from such a strategy would be the detection of universal trends in processual networks and the possibility of identifying a simple ‘alphabet’ of processes. A richer and more explicit set of analytical patterns, approximating intersecting processes, could help evolutionary biologists make better sense of the stunning diversity of evolutionary phenomena, such as early transitions in the evolution of life, the genetic sharing involved in microbial social life, or new joint physiologies, organs, and modes of reproduction involved in evolutionary transitions and in adaptations.

At a more general level, the development of a typology of processes would constitute a genuine attempt toward unification within the (logically pluralistic) biological sciences. To paraphrase Bertalanffy (1968: 48), a unitary conception of the biological world may be based, not upon the possibly futile and certainly far-fetched hope to finally reduce all levels of reality to the level of molecules, but rather on the potential isomorphy of processes in different biological fields.

Acknowledgements

We thank Dan Nicholson, John Dupré, and three anonymous reviewers for comments that greatly improved the manuscript. We also thank Alexander Jaffe for critical revisions of the text. EB thanks the ERC grant FP7/2007-2013 Grant Agreement # 615274 and ANR-13-BSH3-0007 Explabio.

References

Bibliography references:

Abby, S. S., Tannier, E., Gouy, M., and Daubin, V. (2012). Lateral Gene Transfer as a Support for the Tree of Life. *Proceedings of the National Academy of Sciences* 109 (13): 4962–7.

(p.297) Alon, U. (2006). *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton: CRC Press.

Alvarez-Ponce, D., Lopez, P., Bapteste, E., and McInerney, J. O. (2013). Gene Similarity Networks Provide Tools for Understanding Eukaryote Origins and Evolution. *Proceedings of the National Academy of Sciences* 110 (17): E1594–603.

Bapteste, E. (2014). The Origins of Microbial Adaptations: How Introgressive Descent, Egalitarian Evolutionary Transitions and Expanded Kin Selection Shape the Network of Life. *Frontiers in Microbiology* 5 (83).

Bapteste, E. and Burian, R. M. (2010). On the Need for Integrative Phylogenomics, and Some Steps toward Its Creation. *Biology and Philosophy* 25 (4): 711–36.

Bapteste, E. and Dupré, J. (2013). Towards a Processual Microbial Ontology. *Biology and Philosophy* 28 (2): 379–404.

Bapteste, E., et al. (2009). Prokaryotic Evolution and the Tree of Life Are Two Different Things. *Biology Direct* 4, 34.

Bapteste, E., et al. (2012). Evolutionary Analyses of Non-Genealogical Bonds Produced by Introgressive Descent. *Proceedings of the National Academy of Sciences* 109 (45): 18266–72.

Bouchard, F. (2014). L'évolution par sélection naturelle. In T. Hoquet and F. Merlin (eds), *Précis de philosophie de la biologie* (pp. 251–61). Paris: Vuibert.

Braillard, P.-A. (2008). *Enjeux philosophiques de la biologie des systèmes*. PhD dissertation, University of Brussels.

Brucker, R. M. and Bordenstein, S. R. (2012). Speciation by Symbiosis. *Trends in Ecology & Evolution* 27 (8): 443–51.

Claverie, J. M. and Ogata, H. (2009). Ten Good Reasons Not to Exclude Viruses from the Evolutionary Picture. *Nature Reviews Microbiology* 7: 306–11.

Corel, E., Lopez, P., Méheust, R., and Bapteste, E. (2016). Network-Thinking: Graphs to Analyze Microbial Complexity and Evolution. *Trends in Microbiology* 24 (3): 224–37.

Dagan, T. and Martin, W. (2007). Ancestral Genome Sizes Specify the Minimum Rate of Lateral Gene Transfer during Prokaryote Evolution. *Proceedings of the National Academy of Sciences* 104 (3): 870–5.

Doolittle, W. F. (1999). Phylogenetic Classification and the Universal Tree. *Science* 284 (5423): 2124–9.

Duboule, D. and Wilkins, A. S. (1998). The Evolution of 'Bricolage'. *Trends in Genetics* 14 (2): 54–9.

Eigen, M. (1971). Selforganization of Matter and the Evolution of Biological Macromolecules. *Naturwissenschaften* 58 (10): 465–523.

Eigen, M. and Schuster, P. (1977). The hypercycle. A Principle of Natural Self-Organization. Part A: Emergence of the Hypercycle. *Naturwissenschaften* 64 (11): 541–65.

Eigen, M. and Schuster, P. (1981). Comments on 'Growth of a Hypercycle' by King (1981). *Biosystems* 13 (4): 235.

Eigen, M., et al. (1991). The Hypercycle: Coupling of RNA and Protein Biosynthesis in the Infection Cycle of an RNA Bacteriophage. *Biochemistry* 30 (46): 11005–18.

Felsenstein, J. (2004). *Inferring Phylogenies*. Seattle: University of Washington.

Filee, J. (2014). Multiple Occurrences of Giant Virus Core Genes Acquired by Eukaryotic Genomes: The Visible Part of the Iceberg? *Virology* 466–7: 53–9.

Filee, J., Forterre, P., and Laurent, J. (2003). The Role Played by Viruses in the Evolution of Their Hosts: A View Based on Informational Protein Phylogenies. *Research in Microbiology* 154 (4): 237–43.

Filee, J., et al. (2006). A Selective Barrier to Horizontal Gene Transfer in the T4-Type Bacteriophages That Has Preserved a Core Genome with the Viral Replication and Structural Genes. *Molecular Biology and Evolution* 23 (9): 1688–96.

(p.298) Filee, J., Siguier, P., and Chandler, M. (2007). I Am What I Eat and I Eat What I Am: Acquisition of Bacterial Genes by Giant Viruses. *Trends in Genetics* 23 (1): 10–15.

Gilbert, S. F., Sapp, J., and Tauber, A. I. (2012). A symbiotic view of life: we have never been individuals. *Quarterly Review of Biology* 87 (4): 325–41.

Halary, S., et al. (2010). Network Analyses Structure Genetic Diversity in Independent Genetic Worlds. *Proceedings of the National Academy of Sciences* 107 (1): 127–32.

Hatfull, G. F. (2008). Bacteriophage Genomics. *Current Opinion in Microbiology* 11 (5): 447–53.

Hendrix, R. W., et al. (1999). Evolutionary Relationships among Diverse Bacteriophages and Prophages: All the World's a Phage. *Proceedings of the National Academy of Sciences* 96 (5): 2192–7.

Huson, D., Rupp, R., and Scornavacca, C. (2010). *Phylogenetic Networks*. Cambridge: Cambridge University Press.

Jachiet, P. A., et al. (2014). Extensive Gene Remodeling in the Viral World: New Evidence for Nongradual Evolution in the Mobilome Network. *Genome Biology and Evolution* 6 (9): 2195–205.

Keeling, P. J. and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* 9 (8): 605–18.

Kitano, H. (2002). Computational Systems Biology. *Nature* 420 (6912): 206–10.

Ku, C., et al. (2015a). Endosymbiotic Gene Transfer from Prokaryotic Pangenomes: Inherited Chimerism in Eukaryotes. *Proceedings of the National Academy of Sciences* 112 (33): 10139–46.

Ku, C., et al. (2015b). Endosymbiotic Origin and Differential Loss of Eukaryotic Genes. *Nature* 524 (7566): 427–32.

Lapointe, F. J., et al. (2010). Clanistics: A Multi-Level Perspective for Harvesting Unrooted Gene Trees. *Trends in Microbiology* 18 (8): 341–7.

López-García, P., et al. (2015). Bacterial Gene Import and Mesophilic Adaptation in Archaea. *Nature Reviews in Microbiology* 13 (7): 447–56.

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Reidel: Boston.

McDaniel, L. D., et al. (2010). High Frequency of Horizontal Gene Transfer in the Oceans. *Science* 330 (6000): 50.

McInerney, J. O., O'Connell, M. J., and Pisani, D. (2014). The Hybrid Nature of the Eukaryota and a Consilient View of Life on Earth. *Nature Reviews Microbiology* 12 (6): 449–55.

Méheust, R., Lopez, P., and Bapteste, E. (2015). Metabolic Bacterial Genes and the Construction of High-Level Composite Lineages of Life. *Trends in Ecology & Evolution* 30 (3): 127–9.

Milo, R., et al. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298 (5594): 824–7.

Nelson-Sathi, S., et al. (2012). Acquisition of 1,000 Eubacterial Genes Physiologically Transformed a Methanogen at the Origin of Haloarchaea. *Proceedings of the National Academy of Sciences* 109 (50): 20537–42.

Nelson-Sathi, S., et al. (2015). Origins of Major Archaeal Clades Correspond to Gene Acquisitions from Bacteria. *Nature* 517 (7532): 77–80.

Nicholson, D. J. (2012). The Concept of Mechanism in Biology. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 152–63.

Raoult, D. (2009). There Is No Such Thing as a Tree of Life (and of course Viruses Are Out!). *Nature Reviews Microbiology* 7 (8).

Skippington, E. and Ragan, M. A. (2011). Lateral Genetic Transfer and the Construction of genetic exchange communities. *FEMS Microbiology Reviews* 35 (5): 707–35.

Sole, R. V. and Valverde, S. (2006). Are Network Motifs the Spandrels of Cellular Complexity? *Trends in Ecology & Evolution* 21 (8): 419–22.

(p.299) Szathmáry, E. (2015). Toward Major Evolutionary Transitions Theory 2.0. *Proceedings of the National Academy of Sciences* 112 (33): 10104–11.

Villarreal, L. P. and Witzany, G. (2010). Viruses Are Essential Agents within the Roots and Stem of the Tree of Life. *Journal of Theoretical Biology* 262 (4): 698–710.

von Bertalanffy, L. (1968). The Meaning of General System Theory. In L. von Bertalanffy (ed.), *General System Theory: Foundations, Development, Applications* (pp. 30–53). New York: George Braziller.

Wang, M. and Caetano-Anolles, G. (2009). The Evolutionary Mechanics of Domain Organization in Proteomes and the Rise of Modularity in the Protein World. *Structure* 17 (1): 66–78.

Wilkins, A. S. (2007). Between 'Design' and 'Bricolage': Genetic Networks, Levels of Selection, and Adaptive Evolution. *Proceedings of the National Academy of Sciences* 104 (Suppl. 1): 8590–6.

Williams, T. A., et al. (2013). An Archaeal Origin of Eukaryotes Supports Only Two Primary Domains of Life. *Nature* 504 (7479): 231–6.

Yafremava, L. S., et al. (2013). A General Framework of Persistence Strategies for Biological Systems Helps Explain Domains of Life. *Frontiers in Genetics* 4 (16).

Yutin, N., Raoult, D., and Koonin, E. V. (2013). Virophages, Polintons, and Transpovirons: A Complex Evolutionary Network of Diverse Selfish Genetic Elements with Different Reproduction Strategies. *Virology Journal* 10 (158). **(p. 300)**

Notes:

([1]) Furthermore, these donors may themselves already harbour phylogenetically composite genomes, due to LGT between prokaryotes in their own ancestral lineages.

([2]) One of us is currently developing similarity network methods, searching for communities and for simple patterns in gene-sharing networks: see Bapteste et al. 2012 and Alvarez-Ponce et al. 2013. These methods operate under the expanded version of Doolittle's drawing, enhancing the study of intersecting processes in scientific analyses, with promising results.

([3]) In particular, gene externalization may be random and occur at a high rate, which would not be visible from LGT analyses, if the host cell receiving a transferred gene selects against the residency of some of the externalized genes (for example, informational genes may be more externalized than transferred).

([4]) Preliminary analyses suggest that gene externalization is a general property of life: cellular genomes, especially prokaryotic ones, decoalesce; as do viral genomes. They also suggest that this process is likely to have been ongoing for a very long time. This proposition is backed up by the scientific literature on virus evolution by gene accretion (Hatfull 2008; Hendrix et al. 1999; Filee et al. 2007) and on DNA dissemination via gene transfer agents (McDaniel et al. 2010), as well as by the debates regarding the respective origins of viruses and cells (Villareal and Witzany 2010; Filee et al. 2003; Claverie and Ogata 2009).

([5]) Eigen did not tackle this major question of biology with a tree, showing the divergence of DNA or protein, but with evocative sketches and very complex systems of differential equations, which describe the rate of reproduction of the various nucleic and proteic elements of the system in order to predict its behaviour.

([6]) Indeed, systems biologists expect biological networks to feature at least a recurring small set of basic building blocks called ‘network motifs’, which are practically defined as patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks (Milo et al. 2002; Alon 2006).

([7]) By contrast, the information contained in the fragment of DNA that codes for this particular enzyme is not sufficient to produce a ready-made operational intracellular robot. Consequently, a phylogenetic analysis of an enzyme offers an underdetermined argument for the evolution of that enzyme’s function, and should thus be revisited with an approach that incorporates intersecting processes.

([8]) Interestingly, similar views have been expressed by prominent evo-devo scientists such as Duboule and Wilkins (1998). According to these authors, networks explain how internal constraints lead to restrictions in the production of evolutionary novelties.

([9]) For example, the ‘big bang’ in the creation of composite genes has been proposed to have been paralleled by a ‘big bang’ in the evolution of interaction networks during the early history of life on Earth (see Wang and Caetano-Anolles 2009).

([10]) Moreover, however recurrent, such network motives are not necessarily adaptive. They could be the ‘spandrels’ of network complexity; a by-product of network building rules (Sole and Valverde 2006).