

Trashing life's tree

L. R. Franklin-Hall

© Springer Science+Business Media B.V. 2010

Abstract The Tree of Life has traditionally been understood to represent the history of species lineages. However, recently researchers have suggested that it might be better interpreted as representing the history of cellular lineages, sometimes called the Tree of Cells. This paper examines and evaluates reasons offered against this cellular interpretation of the Tree of Life. It argues that some such reasons are bad reasons, based either on a false attribution of essentialism, on a misunderstanding of the problem of lineage identity, or on a limited view of scientific representation. I suggest that debate about the Tree of Cells and other successors to the traditional Tree of Life should be formulated in terms of the purposes these representations may serve. In pursuing this strategy, we see that the Tree of Cells cannot serve one purpose suggested for it: as an explanation for the hierarchical nature of taxonomy. We then explore whether, instead, the tree may play an important role in the dynamic modeling of evolution. As highly-integrated complex systems, cells may influence which lineage components can successfully transfer into them and how they change once integrated. Only if they do in fact have a substantial role to play in this process might the Tree of Cells have some claim to be the Tree of Life.

Keywords The tree of life · Phylogenetics · Evolutionary modeling · Lineages · Tree of cells

Introduction

The Tree of Life (TOL) is most simply seen as a (1) representation of life's history that (2) has a tree structure. But what is *life's history*? And what is a *tree structure*?

L. R. Franklin-Hall (✉)
New York University, New York, NY 10003, USA
e-mail: lrf217@nyu.edu

Biologists have filled out (1) and (2) in different ways. Traditionally, biologists aspired to construct what we can call the ‘simple tree’, a “tree representing genealogical relationships of all known species” (Yates et al. 2004, p. 7). This tree is a representation of *species lineages* using a *tree graph*, that is, one composed of recursively bifurcating branches. In the simple tree, a biological species corresponds to a twig, and a higher clade, composed of an ancestor and all of its descendants, to collections of branches. Every species is placed at one location on the tree, since each species has just one parent species from which it derived.

This simple tree of life has recently been ruled “obsolete” (Puigbò et al. 2009) because it can “not provide an accurate depiction of the processes that have shaped life’s history” (Fournier et al. 2009, p. 2229). Genomic evidence shows that organisms—particularly prokaryotes—receive genes from a variety of sources, sometimes from organisms in *other* species. This happens via entry of foreign DNA into a cell followed by integration of that DNA in the host genome, in a process called lateral gene transfer (LGT) (Planet 2002). Because of LGT, the genealogical histories of different parts of the genome can differ and it is not possible to represent all these different histories using a single, recursively bifurcating graph. Species which have received genes from multiple parent species cannot be positioned on a single branch of the simple TOL.

In light of this challenge to the simple tree, researchers have, broadly speaking, taken up one of two remaining options. One option is to tailor our understanding of the TOL so that it can accurately represent the biological facts (Galtier and Daubin 2008; Ciccarelli et al. 2006; Puigbò et al. 2009). The tree may, for example, be understood to represent the history of particular *core genes*, those which were not transferred laterally between species. Or the ‘tree’ could be, technically speaking, some other kind of graph, including one that allowed reticulation. Let’s call all modifications of the simple tree ‘refined trees,’ which come in different varieties depending on the details. After making the appropriate modifications, proponents of this approach maintain that a Tree of Life *really does exist* (Lienae and DeSalle 2009; Gribaldo and Brochier 2009).

A second option is to reject the *Tree of Life* altogether, be it simple or refined. Ford Doolittle and collaborators have been instrumental in arguing for this alternative (Doolittle 1999; Doolittle and Baptiste 2007; Gogarten et al. 2002). These researchers suggest that frequent inter-species gene transfer does not simply *obscure* the tree of life, but *deconstructs* it. Tree-detractors may admit that refined trees represent *something*—but not enough to have a legitimate claim to the illustrious title, the Tree of Life. After all, some refined trees reflect only a tiny quantity of genetic data, and even optimistically this may be only 1% of the total genomic history of organisms (Dagan and Martin 2006). Detractors urge that it would be wise to abandon this degenerating tree research program and admit that the tree of life *does not exist* (Baptiste and Boucher 2008; Lopez and Baptiste 2009; Lawton 2009).

Which option is preferable? In light of the biological complexities that the genomic era has uncovered, should we continue to *refine* the tree, or *trash* it? This paper aims to evaluate some of the reasons for taking the second course of action. These considerations will prove more complicated than those for which the simple

tree was rejected. The simple tree was rejected because it was *inconsistent* with the evidence. Interpreted as representing full species histories, all simple trees misrepresent those histories.¹ But refined trees have been constructed so that they accommodate the evidence when correctly interpreted, primarily through circumscribing the portion of life's history which the tree is supposed to represent. Consequently, the debate about the viability of refined trees turns on more subtle issues, one of which is the nature of scientific representation.

This paper begins by suggesting that two initially tempting critiques of the refined tree are unpersuasive, in part because they are based on an oversimplified understanding of scientific representation. The first critique concerns the *kinds of entities* that a historical representation needs to track; the second critique concerns how *complete* a representation must be to be adequate. Both of these critiques are found wanting.

Yet even if some challenges to the tree fail, others may not. So, after surveying the two bad reasons to reject the tree, I consider some better ones. These better reasons call attention to the dearth of *purposes* served by a refined tree. Many of the purposes which Darwin, as well as contemporary biologists, hoped a tree of life to serve cannot be discharged by refined trees. This is bad news for tree-lovers. However, failure to serve any particular set of purposes is not definitive, as all representations will fall short by some measure. The paper concludes by helping tree advocates to articulate *one* purpose, understanding evolutionary dynamics, which might justify considering a refined tree to be *The Tree of Life*. In particular, I will suggest that if the cell plays a special role in explaining evolutionary dynamics, the *Tree of Cells*, one version of a refined tree, may in fact be the *Tree of Life* despite extensive LGT.

Although this paper is more sympathetic to refined trees than are many contributors to this special issue (i.e., Doolittle 2010; Baptiste and Burian 2010), its goal is *not* to defend the refined tree. Instead, the task is to examine, clarify, and evaluate some of the reasons offered to abandon it. This analysis of reasons is inevitably partial. Some factors relevant to the debate, for example those concerning our ability to reconstruct the tree of life, are ignored. Others are considered in depth. This partiality reflects a focus on the issues that philosophical analysis may illuminate. It does not necessarily reflect those considerations which are—or should be—dispositive. Because of this partial survey, the refined tree is not, in the end, declared triumphant; however, the grounds for such evaluations are clarified, with some reasons discarded and others embraced.

¹ One issue ignored here is the degree to which even the simple tree may be an *approximately* true description of life's history (Godfrey-Smith 2009). If one could argue that the *simple* tree description were approximately true of life's history, it would neither be necessary to refine the tree nor to discard it. Evaluating this claim is beyond the scope of the present paper and ultimately depends on what purpose we want the representation to serve. Those who emphasize the extent of LGT would likely press that LGT is so important in shaping cellular life that no serious biological purpose could be served by a representation which didn't include it, even one that was in some respects "approximately true".

A refined tree of life: the tree of cells

Before scrutinizing reasons to reject a refined tree of life, we should sketch the refined tree on which the following discussion will focus. That tree is the one that has received the most critical attention: the *Tree of Cells* (TOC) (Baptiste et al. 2005; Doolittle and Baptiste 2007; Doolittle 2009a; Doolittle 2010; Fournier et al. 2009; Zhaxybayeva et al. 2004). Researchers at work constructing a TOL often take their project to be that of uncovering the TOC (Lerat et al. 2003; Puigbò et al. 2009; Woese 2002; Gribaldo and Brochier 2009) or can be interpreted as doing so (Ciccarelli et al. 2006; Philippe and Douady 2008).²

The TOC is a representation of *cell lineages* using a tree graph with *recursively bifurcating branches*. The lineages traced by this tree are *cell* lineages, not species lineages as in the simple tree. These lineages branch only (although not always³) when a cell forms two progeny cells. The TOC would ultimately trace cell lineages back to the last common cellular ancestor(s) of life on earth.

Because prokaryotes are, for the most part, single-celled organisms, the Tree of Cells is sometimes called the Tree of Organisms, or the Organismal Phylogeny (e.g., Zhaxybayeva and Gogarten 2004; Gribaldo and Brochier 2009; Lerat et al. 2003). These are slightly controversial designations, as it isn't clear what exactly an organism is: a functionally integrated unit, a genetically homogenous entity, or something else (Wilson 1999, 2005)? Among prokaryotes, difficulties result from the fact that cells can form complex symbiotic communities which form their own lineages (O'Malley and Dupré 2007; Shapiro and Dworkin 1997; Bouchard 2010). In order to avoid this complex issue, here we focus simply on the Tree of Cells; only if prokaryotic cells are themselves organisms will this also be a tree of organisms.

The TOC is a tempting successor to the simple tree because frequent LGT does not undermine the treeness of cell history. In this regard, the TOC resembles the simple tree. While gene histories are reticulate, with some genes moving from cells in one species to those in another species, the *cell* tree always bifurcates.⁴ This follows from simple facts about cellular reproduction. Cells always come into being

² Some discussions of the TOL do not explicitly state what the tree is intended to represent. Furthermore, examining tree construction methods may not uncover this, as the *signal* used to reconstruct the tree may be different from the plausible representative aim of the tree. To use a simple example, a biologist may conduct a census of a panda population by counting panda droppings. Although she counts droppings, her aim is obviously to measure panda number. Similarly, some biologists constructing trees based on rRNA divergence may aim to construct a more substantive tree of life, not simply an rRNA gene tree. However, the nature of this "more substantive tree" is not always transparent.

³ All cell-division events need not be represented as a bifurcation in the TOC, as a full interpretation of the tree could include some standard (e.g., ecological) for grouping cell lineages together. Important for our purposes here is that any split in those grouped lineages would correspond to some (set) of cell-division events.

⁴ The TOC works most naturally as an interpretation of the tree of prokaryotes, not of the tree of sexually reproducing organisms. This paper focuses on challenges to the TOC from LGT, so it will suffice to see whether the history of prokaryotic life can be identified with the TOC. Although the history of sexually reproducing organisms can less easily be understood in terms of the TOC, the relative lack of LGT in those organisms means that we may be able to maintain a "simple" interpretation of the TOL, in terms of recursively bifurcating species, in that domain.

from other cells through cell division. Thus, the *cellular* history of life can be represented as a recursively bifurcating tree, even if genomic history cannot be.⁵

Looming large for tree-detractors is the fact that the TOC may differ from every particular gene tree when examined over long time-frames. Even if LGT occurred at a relatively low rate, some models indicate that all but a few genes may have been transferred laterally at *some* point in the history of life (Dagan et al. 2008; Dagan and Martin 2007). Consequently, it may be that no one gene type has the same bifurcation history as the TOC.

The difficulty in equating the cell tree with any gene tree is particularly obvious in the early stages of life on earth. The first hereditary material may not have been nucleic acid, but a very different sort of substance, such as clay crystals (Cairns-Smith 1985). If so, a tree of cells couldn't possibly reflect the history of any gene, since the very stuff out of which genes were constituted, and (presumably) the means through which they carried information, has not been stable over the history of life.

However, the possibility of such radical transformations in the genetic material does not undermine the existence of the cell lineages which compose the TOC. Cairns-Smith suggests a helpful metaphor (often attributed in the microbiological literature to Gary Olsen) to illustrate the possibility of such lineages: "A long line of organisms ... is a rope made up, as most ropes are, of long overlapping fibers. It is not necessary that any fiber extends from one end of the rope of the other" (1985, p. 61). The fibers, even those composed of completely different sorts of stuff (not necessarily just nucleic acids), can vary, but a rope extends from end-to-end, held together by the locally overlapping fibers.

A rope metaphor is also helpful in picturing cell lineages after the advent of DNA-heredity. Cell lineages are lineages of cellular ancestry, and they reflect "the majority consensus of genes passed on over short time intervals" (Zhaxybayeva et al. 2004, p. 254). Genes can enter a lineage by LGT without altering what we would identify as the cell lineage, since entering genes don't alter the majority-consensus of genetic inheritance at a particular time. The only cell lineages that are not tree-like are those in which "two organisms make co-equal contributions to a new line of descent" (Zhaxybayeva et al. 2004, p. 254).⁶ Assuming this sort of circumstance is rare or absent, the genealogy of cells in prokaryotes can be accurately represented as a *tree*, one we are calling the Tree of Cells (TOC).

Again, this paper will say nothing about how to reconstruct the TOC, which may or may not be possible depending on whether high rates of LGT obscure the vertical history of cells. There is something to be gained by ignoring the many

⁵ Certain phenomena, such as endosymbiosis, can make even cellular history non-tree-like if two cells contribute equally to a progeny cell. For our purposes what matters is that a history of cells would still be dramatically *more* tree-like than would be full genomic histories, because the objections I consider against the TOC are based on the existence of LGT. Furthermore, endosymbiosis disrupts the cell tree primarily among eukaryotes, which are not the focus of this paper.

⁶ It has been suggested that cell division may involve the transmission of "cytoplasmic structure" which can take different forms, contributing to the importance of the cellular lineage over particular gene lineages (Thaler 2009). If there is such heritable cytoplasmic structure, this would simply add one more kind of "fiber" to the rope that characterizes the cellular lineage.

epistemological and empirical issues and focusing discussion on whether this conception of the refined tree—the TOC—is even a *coherent* successor to the simple tree.

Bad reasons to trash the tree

Now that we have described one version of the refined tree of life, the TOC, we will evaluate a few reasons to reject it. Some common reasons for rejecting this version of the tree will prove unpersuasive: some because they assume an overly reductive conception of the TOC; others because they rest on an overly simplified understanding of scientific representation. The section following considers more persuasive reasons to reject the TOC as a conception of the TOL.

Leery about lineages

As described above, cell lineages in species with frequent LGT are analogous to ropes: there is local overlap between fibers (or gene-lineages) but there may not be a single fiber (or gene-lineage) which extends from end to end. When many of the fibers are short and more genetic variation enters the lineage through LGT than through endogenous (point) mutation, we can call a lineage “open” (Boucher and Baptiste 2009). There is controversy about how many lineages are open and to what degree. Since the aim here is to criticize reasons offered against the tree, the argument will be strongest if we assume the worst for the TOC advocate. Consequently, let us assume, contrary to fact, that all lineages are open.

Some biologists have questioned whether what have been called “open lineages” really qualify as *lineages* at all. Some claim that the “integrity” of cell (and organism) lineages has been violated by LGT (Doolittle 1996, p. 8799). Others worry whether LGT could “render the concept [...] of the organismal phylogeny impossible” (Philippe and Douady 2003, p. 498) and ask whether “the replacement of every gene in a genome invalidates the tree concept for organism lineages” (Zhaxybayeva et al. 2004, p. 254). This section critically describes two sorts of concerns one might have with cell (aka organism) lineages—first that they are implicitly essentialist, and second that, if non-essentialist, they fall prey to a metaphysical paradox. Both concerns I find to be wanting.

First concern: essentialism

In order to examine our first concern with the TOC, it is worth reviewing an equivalent charge against other kinds of refined trees. Other versions of the TOL have been rejected for being “essentialist”. The ribosomal RNA tree in particular has been subjected to this critique.⁷ The rRNA tree tracks *particular kinds* of genes through time, genes coding for ribosomal RNA. Critics have asked why a particular gene or gene suite should “define” life’s history. Doing so seems either essentialist

⁷ Recently, Doolittle (2009b) has coined the term “riboessentialist” in this connection.

or arbitrary and unmotivated (Baptiste and Boucher 2008; Doolittle 1996; Doolittle 1999; Franklin 2007). Good post-Darwinian biologists should have abandoned all forms of essentialism: no one part of the cell is more essential (metaphysically) than any other. Thus, the ribosomal RNA tree is not the TOL.

The same argument can be run against the TOC by claiming that it too is implicitly essentialist. What 'essential' feature might the TOC represent? Doolittle has suggested that the TOC really tracks lineages of the cell envelope (cell membranes plus the cell wall). After all, as we've emphasized, the TOC doesn't track the passage of any *gene type* over time, since no "fibers" extend the length of the lineage. Consequently, it may appear that the only thing that the TOC *could* be representing is the entity that *contains* those genes—the cell envelope. It makes sense then to question why the envelope's fate is so important. Why should lineages of cell envelopes constitute the TOL?

If the TOL actually did aim to simply track the envelope history, this argument would succeed. However, this is not the most promising interpretation of the TOC. A lineage of cells need not be identical to a lineage of any particular part of the cell, whether gene, membrane, or otherwise. The cell is a complex machine with *lots* of parts (genes, membranes, proteins, etc.), and lots of organizational features relating these parts (gene order on chromosomes, metabolic networks, etc.). As I characterized the TOC above, lineages in the TOC trace the consensus of parts of the cell locally. Due to facts about cellular reproduction, this will, for the most part, lead the TOC to be a representation of the tree of cellular reproduction, and it will for the most part track the cell envelope. But this isn't what it aims to represent, which is the history of a whole host of different features, including genes, membranes, and their interrelations. Once we see this, the essentialist worry evaporates.

The temptation to identify the TOC with a lineage of one of its parts may be a remnant of the replicator view of evolution, as pressed most influentially by Richard Dawkins. On such a view, certain "agents" (like genes) are considered to be the entities driving evolution, the divergence of which the tree of life traces. Crucially, these entities must be capable of perfectly replicating themselves. The only candidates for such a role are genes or membranes.

But philosophers have increasingly questioned the replicator framework (Sterelny and Griffiths 1999). Evolution simply requires resemblance between parent and child, not perfect replication of any part (Godfrey-Smith 2009). It is beyond the scope of this paper to argue against the replicator framework. But those who have abandoned it should not be misled into the reduction of the TOC to the tree of cell envelopes.

Second concern: metaphysical paradox

The previous section suggested that there is no special cellular feature—not even the cell envelope—which lineages in the TOC aim to represent. Consequently, the TOC shouldn't be taken to be clandestinely essentialist. But perhaps this opens the TOC up to a second concern. The problem is not that cell lineages take some feature as *essential*, but that they *do not* take anything as essential: they do not map

anything in particular over the history of life. At different points in life's history, lineages may depict the transfer of different hereditary materials.

To illustrate the alleged difficulty, Zhaxybayeva et al. (2004) and Doolittle (2004, 2005) invoke the famed Ship of Theseus, as do other papers in this issue (e.g., Beiko 2010; Andam et al. 2010). Imagine a ship preserved over a long time by assiduously replacing one plank after another. Eventually every plank in the ship is replaced.⁸ Now one may wonder whether the old ship still exists, though none of the original matter remains, or whether the old ship has simply ceased to exist. The ship at different points in time is supposed to be equivalent to different cells along a lineage in the TOC. At one point, the lineage's cell is composed of one set of gene types or "planks". At another point, it is composed of a different set of genes types or "planks". Writers invoking the Ship of Theseus paradox think we would do best to track the movement of the planks or gene types and give up on the elusive ship.⁹

However, the paradox of the Ship of Theseus scenario is irrelevant to the problem of cell lineages.¹⁰ What is at issue in debates about the TOL is the existence of a *lineage* constituted by different entities at different times, not whether one *cell* can survive radical genetic replacement. And for a *lineage* to exist between differently constituted cells simply requires the local continuity and majority consensus in the transfer of parts. Whatever you may say about the identity of the ship through the change, there is clearly a set of ship-slices, at different time points, which are closely associated. In order for there to be a puzzle about the lineage, half the planks constituting a boat would have to be added to a boat at one time—not a situation characteristic of LGT.

What has gone wrong is that a metaphysical puzzle that probes the identity and persistence of *entities* has been applied to a slightly different problem: the identity of *lineages*. But these are, metaphysically speaking, two very different beasts. What's a puzzle for one is not a puzzle for the other. As we'll see below, there *are* substantive reasons to doubt whether the TOC has any claim to be the TOL. However, we should not be moved by the wrong reasons. The lineages in the TOC are neither necessarily essentialist nor metaphysically paradoxical.

Apprehensive about representations

Even if cell lineages which compose our proposed refined tree are not themselves problematic, there is a broader concern with the TOC: its incompleteness. The TOC leaves out much of life's history, most importantly the many instances of gene

⁸ Doolittle (2004) pursues the metaphor further, suggesting that the old planks are then used to reconstruct a second ship. This does make the Ship of Theseus paradox more puzzling: Is the second, rebuilt ship really the original Ship of Theseus? But this particularly perplexing aspect of the problem has no close biological analogue – there is no other lineage in which all the parts of the first lineage are progressively integrated, eventually forming a genetically identical organism. Rather the "old planks" are incorporated into many different organisms.

⁹ In contrast to its application here, the Ship of Theseus case is sometimes used to suggest that part substitution does *not* undermine ship identity (Danchin 2002).

¹⁰ Although queasiness about these lineages seems pervasive, I don't mean to suggest that everyone shares it. Boucher and Baptiste, for example, are clear that even after pervasive transfers "the lineages still exist" (2009, p. 532).

transfer between cell lineages. Because of these omissions, biologists conclude that “the history of life cannot be properly represented as a tree” (Doolittle 1999, 2124). Tree-detractors typically emphasize the limitations of a tree-representation and suggest that a more extensive representation would be preferable (Dagan and Martin 2007)—and the more extensive the representation, the better.

We will see here that the incompleteness of the refined tree does not itself provide persuasive reason to discard it. First, because every scientific representation is similarly partial. And second because on some analyses, partial representations are not merely tolerable, but superior. This last claim is a substantive one in the philosophy of science which cannot be fully argued for here, but I will motivate it before drawing conclusions about partial representation in the TOC.

Philosophers have found it fruitful to explore the nature of scientific representation—such as the representation of life’s history—by looking at a simple example: the cartographic map (Toulmin 1953; Kitcher 2001; Godfrey-Smith 2009). Although simpler than many scientific representations, maps are illustrative of at least three features characteristic of all scientific representations.

The first important feature is that there can be different maps representing different aspects of the same domain, and yet each map may be a good one. For example, a Manhattan subway map might depict the topological connections between subway stops, while a waterways map might depict the geometrical locations of islands and canals. These maps depict different kinds of things in the same area and have different standards of accuracy.

A second feature of note is that maps of a domain are typically incomplete. The subway map and the waterways map each represent only some features in the domain of interest. Sometimes maps can be combined, but only a “map” which was a perfect physical duplicate of the island of Manhattan could actually include all the information about the island (Kitcher 2001).¹¹

A third feature is that, even when it is possible to create a more extensive map, it isn’t always better to do so. Adding more features to a representation will not necessarily aid the purposes at hand. Scientific modelers—a species of map-makers—even emphasize that they aim to eliminate all extraneous parameters in their models (Forgacs and Newman 2005).¹²

For a homely illustration of this last point, picture a small pond in a city park. A person who wanted to predict where her toy boat would go upon putting it in the water would normally need a very detailed map representing every inlet and peninsula and perhaps even depicting the locations of the surrounding buildings that affect wind currents. But some ponds are structured so that a more limited representation would be preferable. In particular, suppose the pond had two drains at the east and west ends of the pond. And suppose that, because of the drains’ influence on the water circulation, a boat placed at one end of the pond always ends up floating above the drain on that same side. For our purposes, a good map of *this*

¹¹ This is assuming that the system is completely non-redundant, meaning that no aspect of its structure can be inferred from any other. Only in that case is it its own simplest description (Simon 1996).

¹² For different defenses of the value of prescinding from the details in explanatory contexts, see Batterman (2002), Kitcher (1999), and Strevens (2008).

particular pond would best omit many features of the system included in a map of a pond without such drains. The only feature relevant to a boat's final destination is whether the boat is initially put into the water on the eastern or western end of the pond; the surrounding environment and the precise shape of the coastline are irrelevant. Including such extraneous information diminishes the value of the map for the person guided by an interest in knowing where her boat will go, as it falsely suggests that certain irrelevant features are actually relevant to the boat's trajectory.

In short, sometimes we may have more extensive information about a domain, but the right map may still leave some of that information out. Along these lines, James Watson is thought to have said, uncharitably perhaps, that naturalists were but "stamp collectors" (Wilson 1994, p. 219)—people who indiscriminately collected curiosities, not to reveal anything in particular about the way the world works, but simply to gather together everything that had attracted their attention.¹³ Even if your first impulse is to defend the naturalist, Watson's remark has some value. Scientific representations aim to reveal something underlying the myriad of details. You only include all the details if you cannot distinguish between the relevant and the irrelevant.

Taken together, the three above considerations provide a framework in which we can defuse possible concerns about the limitations of the refined tree of life, including the TOC. Following on the first feature, we may admit that there are other maps of the same domain which account for different features. For example, scientists are interested in producing maps that unravel "the complex history of genes and genomes," (Gogarten et al. 2002, p. 2234). Such a map may not depict cell lineages at all. This gene map and the TOC would then be but two representations of the same domain, each depicting some features and leaving out others. Those who believe that there is *one best* representation of all of life would presumably resist this move, but we see here that such resistance is unmotivated.

Following on the second feature, we may point out that in leaving some information out, the TOL is no different from other maps, including the more inclusive frameworks sometimes offered as successors to the TOL, such as a web of life or the net of life. A web-like representation of life's history that included cell history and gene history would still leave out other aspects of the evolving system. For example, such a map might leave out information about extra-reproductive membrane transfer. It might also leave out information about the ecological relationships between organisms which determined the course of evolutionary change, or the non-heritable changes of organisms over their lifetimes which affect their reproductive output. But these omissions, characteristic as they are of scientific representations, are not themselves worrisome, either for the TOC or for some version of the web of life.

Following on the third feature, the TOC may be among the maps which would be *better* for having left out some information, such as particular gene histories. From what we have said so far, however, this is simply a conceptual possibility. How can we decide whether the TOC is, in fact, a map which is good for being limited, for

¹³ For an interesting history of the "stamp collector" jibe, see Johnson (2007).

having ignored some apparently important biological patterns, such as LGT? This will depend on the standards by which we judge maps to be good.

One plausible account is that a good map is one that helps us achieve our aims. Given a set of aims, some maps will aid us in achieving them, others not. As suggested above, sometimes our aims are undermined by including too much information, sometimes by including too little. Just what is essential and what extraneous depends on what these aims are.¹⁴ Thus, if we are to finally evaluate the status of the TOC—and whether, despite its limitations, it has any claim to the title of *The Tree of Life*, we will have to pursue in more depth the aims we actually have for developing representations of life's history.

Good reasons to trash the tree

We just suggested that some scientific maps *are* good even if they represent only some features in their domains, assuming those features necessary to achieve our aims are among them. It appeared at least conceptually possible that the TOC is such a map. Yet this involved a large promissory note: the claim that the TOC *can actually* serve important scientific purposes. Biologists discussing the TOC often suggest as much (Ciccarelli et al. 2006; Lerat et al. 2003; Lake et al. 2003), but little work has been done to characterize these purposes or aims. Doing so will be the central task in what follows. If the TOC can serve important purposes, the TOC may have at least *some* claim to being the TOL. And if it does not, the tree should be trashed.

Darwin's purposes: explaining the nested taxonomic hierarchy

In searching for purposes which the TOC might serve, we will first look to history. One possibility is that the TOC could serve the same purposes that Darwin thought his “great tree” would serve, a tree of bifurcating species lineages called above the ‘simple tree’. Some have argued that Darwin's tree was special because it answered a particular explanatory challenge: to account for similarities and differences between organisms (Doolittle and Baptiste 2007; Coyne 2009). It had long been clear that organisms could be roughly categorized in terms of a hierarchical system of groups, such as species, genus, family, etc. But why? If each species was separately created, it would be surprising that organisms could be so systematized (but see Doolittle 2010).

Darwin suggested that descent with modification, through which different lineages split and diverged over the history of life, was the explanans to account for the pattern of similarities and differences that had been described by systematists,

¹⁴ Metaphysically, there are more pragmatic and more realist ways of thinking about these aims, but we need not wade into such deep waters here. Very briefly though, a realist may think that nature “sets the agenda” for our inquiries, and that the Tree of Life would be real if it were helpful in carrying out that agenda. A pragmatist may deny that there is any naturally given agenda, and would judge the Tree (or other scientific representation) simply in terms of how it furthers our (possibly idiosyncratic) purposes. But on either approach, the TOC would be judged by its ability to help us carry out some scientific agenda.

the explanandum. Two species may be categorized in one genus because they share a number of phenotypic traits, which was then explained through common ancestry: “the characters which naturalists consider as showing true affinity between any two or more species, are those which have been inherited from a common parent, and, in so far, all true classification is genealogical” (Darwin 1859, p. 420). Without Darwin’s great tree, there was no satisfying explanation for such shared traits. After it became clear that all life was related, such explanations became straightforward. And while common ancestry provided explanations for similarities, the affects of natural selection on divergent lineages could explain differences.

Doolittle and Baptiste (2007) have argued that the TOC sketched above will not provide this sort of explanatory pay-off, at least for organisms with frequent LGT. When there has been pervasive LGT, there is no guarantee that the phenotypic traits shared by species in a genus will have come from a common cellular ancestor. Organisms in two species might have traits in common not because they had a recent *cellular* ancestor, but because they had a recent *genetic* ancestor. Interestingly, some of the traits which have been genera-defining in bacteria, such as the presence of a flagellum, are traits whose range can be explained in this way (Baptiste and Boucher 2008). Because of this pervasive problem, Doolittle and Baptiste conclude that “If the tree of cells is taken as a biological fact, it is in any case not the same fact that Darwin accepted as the explanandum of his theory” (2007, p. 6).

Does this mean that the TOC is not the TOL? No, but this is a strike against it. As we discussed above, once we see the TOC as a map-like representation, we will judge the tree based on whether it can help us achieve our scientific aims. If the TOC can serve important aims, it may still be apt to call it the *Tree of Life* despite LGT. Yet here we see that LGT prevents the TOC from performing one task that the simple tree could perform—explaining the hierarchical character of taxonomy. Nevertheless, it would be overhasty to reject the refined tree just because it doesn’t serve this particular purpose. We need to ask whether there are any other important purposes for which the TOC might be more suited.

Explaining evolution: dynamic modeling

As we’ve seen, if the TOC is to have any claim to being the TOL, it needs to serve some important purpose, and yet it cannot deliver on the purpose Darwin is said to have proposed for the TOL. This section considers an alternative perspective on the TOL, and suggests a purpose to which the TOC might be put. What follows is offered in an exploratory spirit as part of an attempt to see what *might* lie behind suggestions and intuitions that the TOC is special, that it should be the central framework for representing the history of (prokaryotic) life. Once again, my goal is not to vindicate the TOC or the TOL.

Much of the discussion of the TOL, and of the inadequacy of the TOC in particular, relates to issues concerning *systematization*. Researchers question the wisdom of calling a particular organism an archaeon when many of its genes are of bacterial origin (Doolittle 1999). It does not seem to make sense to categorize whole organisms based on their histories when the different parts of those organisms have

different histories. Why then, in the face of this problem, do biologists persist in trying to work out a Tree of Life? One explanation is essentially dismissive and debunking. Doolittle suggests that the suitability of tree representations, and the importance of constructing them, is simply something that “we are predisposed to believe by 2000 years of essentialist philosophy. Our attempts to reconstruct [the TOL]—despite evidence that the great majority of genes in the (mostly prokaryotic) history evolved in a web-like fashion—seems to testify to the strength of that predisposition” (Doolittle 2009b, p. 7). A more charitable suggestion, the one developed here, is that while the TOC may not be useful as a framework for biological systematics, it may play a special role in a theory of “evolutionary dynamics,” that is in a theory that can explain how populations of biological entities—including genes—have changed, and will continue to change, through time (Michod 1999; Nowak 2006).

How can we explain the changes that populations of cells and genes have (and have not) undergone and the adaptations they have (and have not) evolved? Answering this question is enormously complex, but here is how we might think about representing part of this process. Following from our earlier discussion of scientific representation, consider constructing a model of a biological system at a given time in an environment. Such a model may be used to predict changes in that system, such as changes in the representation of different traits in a population, perhaps using evolutionary game theory. The model might need to include extensive information, both about the distribution of traits in a population and the structure of that population.

The important questions for us are these: Is it likely that *cells* would have a special place in this model? For example, would it be important to look at populations of cells and diversity within them? And as we traced evolutionary changes through time, would the genealogies of cells have any special status? The hunch of some TOC advocates may be that the answer is “yes.” Cells will be important in explaining the dynamics of evolutionary processes, and explaining such dynamics is an important aim of biology.

To evaluate this suggestion, we can begin by noting some reasons that the cell appears to have a special place in the history of life on earth. Peter Godfrey-Smith suggests that the cell is the distinctive occupant of the role of ‘simple reproducers’, entities capable of reproducing themselves “under their own steam,” and whose parts do not have that capacity, but can only do so through the collective activity of the whole:

The category of simple reproducers is, of course, a pivotal one. On earth, cells are the distinctive occupants of this role, at least at the present time. Higher-level reproduction (reproduction of things like us, bee colonies, and buffalo herds) is elaborately organized cell division, combined with occasional cell fusion. Lower-level reproduction (reproduction of genes and chromosomes, especially) is organized, orchestrated, and made possible by cell division and cell fusion. If a Martian biologist came down to earth, and started afresh on evolutionary theory using none of our usual concepts, I think that pair of facts would loom large. Cells occupy a special place. (2009, pp. 90–91)

Parts of the cell cannot reproduce themselves on their own because reproduction is a complex, coordinated activity, requiring many different parts of the cell to work together. An individual part of the cell, such as the bacterial chromosome, cannot replicate itself. Rather it is replicated via the activities of a variety of systems in the cell—the cell takes in certain nutrients, produces enzymes capable of synthesizing new nucleotides, and eventually divides into two. Reproduction may be a particularly complex task, but it is just one of a variety of activities of the cell (e.g., growth, metabolism) which it can only undertake as a whole and which the parts of the cell cannot execute individually, even when those parts do form lineages of some kind (e.g., genes, membranes).

Why are these interdependencies important for evaluating the TOC? Most simply, interdependencies have a major impact on how cellular systems, and their genes, can evolve. If there are complex interconnections between parts of a system—if the system is not “nearly decomposable”(Simon 1996)—then one part cannot change without affecting the functioning of other parts (Raff 1996). If modifying one aspect of the cell has promiscuous consequences for other aspects of the cell, then this will have two effects. First, cellular interdependencies will influence how genes that are parts of a cellular lineage change through time. And second, such interdependencies will influence which genes can successfully enter a cellular lineage (through LGT). The second of these two effects is crucial for the TOC advocate who is trying to show that the cell—and its lineage—is *particularly* important: it has a substantial influence on the transfer of genes between lineages.

To flesh out this suggestion, distinguish two stages required for what we can call ‘successful LGT’. First, a gene must physically enter a cell and integrate into the host chromosome. Second, the modified cell must be viable, and thus be able to persist and reproduce, in a given environment. How important is the whole cell and its organization in explaining the likelihood of successful LGT? Although there is some evidence that the cell is important for both, let us focus on the second stage, which requires that a recipient cell remain viable. Trivially, this is only possible if the transferred element leads the cell to do something physiologically compatible with the cell’s other activities (or has no affect on those activities). Barring this, the recipient of the transfer would be eliminated, and the transfer would be unsuccessful. Assuming, for ease of exposition, that each gene had an equal chance of *entering* a cell, here are some ways these complex interdependencies might hypothetically affect successful LGT:

- Transferred elements that act alone, requiring no partners to underpin adaptive functional capacities, will, other things being equal, be more likely to be successfully transferred. If so, transfer of whole operons (groups of adjacent, coexpressed and coregulated genes that encode functionally interacting proteins) will be more likely; transfer of elements which don’t act alone will be less likely, whether these be individual genes or gene complexes.
- Transferred elements that cause cellular changes which disable other cellular systems will, other things being equal, be less likely to be successfully transferred. This is because other cellular systems are likely necessary for cellular viability, and disabling those systems will incapacitate the host cell.

- Transferred elements that are very similar in different organisms will, other things being equal, be neither more nor less likely to be successfully transferred. This is because genes very similar to those found in the host cell will presumably do things compatible with, but not adding to other cellular functions. Trivially, a gene identical to that found in the cell could be transferred without detriment to the host.¹⁵

If the cell is indeed a highly integrated machine, these schema (or other like them) should be widely applicable. There should be ways in which features of the whole cell, in the context of a selective environment, guide evolution, both endogenous changes (which we have not explored here) and those through LGT. This might justify a special place for the cell lineage—and trees composed of such lineages—in describing the history of life on earth.

Discussions of LGT and the tree of life have considered a related point under the name the ‘complexity hypothesis’ (Jain et al. 1999; Lake et al. 2005). This hypothesis purports to explain why some genes, particularly informational genes, have been transferred at lower rates than other genes, such as pathogenicity genes. The idea is that the translation machinery is so complex and its function so tightly scrutinized by natural selection that its individual components (ribosomal proteins and certain ribosomal RNAs) cannot function in foreign cytoplasmic contexts, where all the coevolved molecular interactions would be differently evolved. The complexity hypothesis may or may not be correct. Whether it is or not, it is simply one instance of the more general schemas above and illustrates one way in which the cell can have an impact on which genes can enter it—sometimes prohibiting entry, other times allowing certain kinds to occur.¹⁶

Yet it is hard to judge the relevance of these general schema. One way of proceeding is to think about whether there are what we might call ‘dynamic principles,’ perhaps more specific than the above schema though of the same character, which characterize the course of cellular evolution. They would reflect the ways that one part of the cell and its features might affect the other parts. Unfortunately, given the conjectural nature of this exploration, it isn’t easy to point to any such principles presently. Instead, let’s consider an illustration from another science which, like evolution, traces lineages through time, but in which dynamic principles have been characterized. Historical linguistics aims to map and explain

¹⁵ It has been suggested that immediate retention of genes through LGT requires not only neutral effects, but positive selection (Ochman et al. 2000). If so, this principle would have to be slightly modified.

¹⁶ Carl Woese (2002, 2005) has also emphasized the importance of the cell and its organization in determining the likelihood of LGT. He writes that it is “the degree to which (and the way in which) the various componentry is integrated into a cell, the cell design, that determines what is and what is not horizontally transferred” (2005, p. 106). However, when it comes to describing this “cell design” and “organization” Woese focuses on a certain “core” of genes which he says constitutes cellular organization: the “design-defining genes” (2005, p. 111). He suggests that the organismal genealogy—and the TOC—is defined by those particular genes. This account of organismal lineages does fall prey to the essentialist critiques discussed in the previous section. It is counterintuitive to identify the organization of the cell with particular genes, since the total functioning of the cell is not determined by them alone. The enlightened TOC advocate should agree with Woese that cellular features control aspects of LGT, but should not reduce cellular organization to what Woese calls “a small cadre of genes” (2005, p. 111).

linguistic change by explaining changes in grammar, sounds, and lexicon. Languages can be understood in terms of these three features. Just as genes can be transferred between organisms, words, grammatical elements, and sounds can be transferred laterally between previously unrelated languages (e.g., the English adoption of the Polynesian word “taboo”).

Consequently, linguists also ask (in parallel to our question about the cell) whether properties of the whole *language*, such as the combinations of sounds found within it, play any role in explaining language change, or whether one can understand change in terms of independent words where “every word has its own history” (Malkiel 1967). While the debate is ongoing (see Phillips 2006), there are certainly contexts in which languages do play such a role, and in which we can only explain changes through attending to interdependencies between the parts of the language. Here are two examples among many:

- (1) Lexical borrowing: Borrowing of foreign words is influenced by the absence of a synonym in the adopting language. In particular, a language is less likely to adopt a foreign word if there already exists a synonymous word in the language. Based on this principle, we can explain the high number of technological words, the low number of every-day words, which enter languages from other sources (McMahon 1994).¹⁷
- (2) Sound change: Change in one phoneme in a language affects other phonemes in the language. In particular, there is a general tendency for phonemes to stay distinct, so if one phoneme starts to change in a way that pushes it closer to another, the second one will have a tendency to change in response. Based on this principle, we can explain the “Great Vowel Shift” in middle English, where long A displaced long E, which then displaced long I (Labov 1994).

Both of these phenomena have interesting consequences for how languages change over time, and both show the importance of the whole language in explaining such change. As per the lexical borrowing principle, if I want to explain either why a word did or did not enter a foreign language, I need to consider the set of words in that language which have a similar meaning as the potential adoptee. If there are no synonymous words, it will be more likely that the adoption will happen. As per the sound change principle, if I want to explain why a particular word’s sounds changed, I would need to attend to other words in the language and how their pronunciations had shifted. If the pronunciation of other words has shifted, this will affect the likelihood that the focal word will change as well.

In the case of language change, as with the cell, the dynamic principles of the evolving system are not presently well understood.¹⁸ However, to the extent that there are dynamic principles governing sound change which do make reference to whole languages, tying the fate of one part of the language to the fate of another part, there is some reason to consider the language lineage as the primary one that

¹⁷ It is sometimes suggested that word adoption rates are also affected by how similar the sounds of the potential adoptee are to the sounds in the adopting language. This is another dynamic principle in which the language governs which words enter it (McMahon 1994).

¹⁸ For example, the cause of the “Great Vowel Shift” has not infrequently been called “mysterious” (Pinker 1994, p. 252).

historical linguists should trace.¹⁹ In the view of the TOC advocate, this may parallel the cellular case.²⁰

Of course, the presence of just a few token principles would not be significant. The more extensively these principles apply to cells, the larger the explanatory role the cell will have in accounting for evolutionary change. At this point, it is not clear how many or how few such interconnections there are. Still, our general picture of the cell, as explained above, is as a complex, interdependent whole. Functions required for the viability of cell-reproduction, nutrition, and growth—depend on the coordinated action of the parts. Consequently, it would be surprising if there weren't many such principles.

We can anticipate two kinds of objections. One objection would argue that the story told here—an admittedly conjectural one—is empirically wrong about the dynamics of evolutionary systems. Perhaps there are no—or very few—cell-level principles which affect evolution and, by proxy, the evolution of the genes. Here are two reasons for doubting the existence of such generalizations. First, both biologists and philosophers have emphasized the *contingency* of evolution (Gould 1989; Beatty 1995), and in light of that contingency, one might doubt whether there are any such general laws to be discovered.²¹ Second, there is evidence that many organisms—especially prokaryotes—are modular (Beldade et al. 2002). That is, they are constituted by functionally or developmentally independent parts. If *cells* turn out to be particularly modular as well, then there may be no complex interdependencies at the cellular level which underpin such dynamic laws. Thus, the whole cell would not have any special role in accounting for evolutionary change. With respect to both of these considerations, the empirical issues are complex. It suffices to say that if it turns out that there are no substantial dynamic principles at the cellular level, then the argument for the TOC suggested here would fail.

There is a second, more conceptual objection. Even if everything suggested here about the cell's role in explaining evolutionary dynamics were true, there is still room to deny the TOC's claim to being the *Tree of Life*. There are many patterns in nature and the complete account of evolutionary dynamics will have to add to the TOC, at the very least, a complex filigree of web-like gene lineages. By making this

¹⁹ The terrain is too extensive to cover here, but there is no reason to think that the existence of cell or language-level principles has deeply emergentist implications. Following Loewer (2009) let us take emergentists to be those who deny that the laws of physics are dynamically complete. If they were incomplete, there might be nomological structure in the special sciences over and above that determined by the laws of physics. The principles I refer to here should not be identified with this sort of emergent law. Instead, they characterize dependence relations between the parts of a complex system. Some complex systems are organized such that the parts of the system do not have substantial effects on one another; others are tightly interconnected (Simon 1996). The existence of such principles would simply reflect the fact that cells and languages are instances of these highly interconnected systems.

²⁰ In the linguistic case, principles characterize change in a single language through time, while in the cellular case principles apply to changes in lineages. To appreciate the parallels, we can liken time-slices of a language to individual cells in a lineage.

²¹ Ultimately, a certain level of contingency is compatible with the existence of many interesting cell-level generalizations governing dynamics. What matters for the explanatory relevance of cell-level generalizations is that they rule out some possibilities, not that they rule out all but one (leaving no contingency).

addition, however, the tree-like structure may lose its primacy. We will no longer have even a refined tree of life.

There is no denying that gene lineages are important in accounting for some aspects of evolutionary dynamics. The enlightened TOC advocate need not deny this. There will be some features of life's history that the TOC will not shed light on. For example, lateral transfers are important in explaining how organisms can infiltrate new niches (Hacker and Kaper 2000). Because of this it makes sense to be a pluralist about patterns descriptive of life's history.²² But there remain meaningful questions about the relative significance of these patterns. If cell lineages are important in accounting for many aspects of evolutionary dynamics, as the TOC advocate must believe, it could have some sort of explanatory priority over other patterns in nature, in the same way that our simple pond map was preferred to the detailed map in the boat example above. It would be a mistake to think that commitments to pluralism and the rejection of essentialism require denying such a possibility.

Revisiting the 'core'

The complex interdependencies between parts of the cell, and the way that this can affect, among other things, the viability and adaptiveness of organisms undergoing LGT, may give cellular lineages some special status. If so, a representation which tracks divergence in these lineages—the TOC—may have *some* claim to be the TOL. In closing, it is worth contrasting this view with a more standard framing of the debate about the existence of the TOL.

The TOL debate is often framed in terms of the existence and size of a “core” of genes which has never been transferred (Jain et al. 1999; Doolittle 2005; Dagan and Martin 2007). On some views, the core is not simply useful because it makes it easier to uncover the history of cellular lineages, but because the core genes are “essential” for the organism (Lerat et al. 2003). Those who argue that there is a tree of life say that there is such a core, and that LGT is not the “dominant force” in prokaryotic evolution (Lienae and DeSalle 2009, p. 4). Those arguing that there is no real tree of life emphasize the quantity of LGT and deny the existence of a substantial core (Doolittle and Baptiste 2007).

The TOC proponent, on my interpretation, does not put so much emphasis either on the existence of a core or on the frequency of LGT. This is because the presence

²² There may well be some less-enlightened TOL advocates who would deny even this, claiming that there is one unified, *tree* representation of life's history which accounts for *everything*. As should be obvious, this view is too strong. However, it is not clear that there are many contemporary advocates of this strong view. Even many advocating something they call the tree of life say that there exist multiple representations of life's history: “a web of life and a tree of Life are both useful representations of two different biological processes...A web of life illustrates the HGT component of genome evolution, whereas a tree of Life illustrates the history of cells” (Gribaldo and Brochier 2009, p. 515). Alternately, some tree advocates claim that “the organismal [cellular] phylogeny provides the backdrop” (Lerat et al. 2003, p. 101) on which we might represent LGT. As these tree-advocates suggest representing both cellular and genetic lineages, what is at stake is the relative importance of these patterns and whether the cell tree is a particularly interesting and explanatory one.

or absence of a core is actually orthogonal to the question of whether the cell lineage tree has claim to being the TOL. What matters instead is the degree to which there are complex interdependencies between the parts of the cell that influence the course of evolution—sometimes this may promote lateral gene transfer of one kind, other times it may hinder it, but in various ways these interdependencies direct it.

These comments on the nature of the TOC are tentative. The history of cellular lineages might end up either more or less important in the map of the total history of life than I've suggested here. More important than any particular conjecture, the central aim of this paper has been to distinguish the good *sorts* of reasons to reject the TOL from the bad. This discussion has urged that the debate about the existence and reality of the Tree of Life should be evaluated, not in terms of overly-metaphysical concerns about essentialism nor on the basis of an oversimplified view of scientific representation, but instead in terms of the purposes which such representations will serve. At this point, there should not be overwhelming optimism that the TOC—or some other refined tree—will be vindicated by this test. But if it is to be rejected, let it be for this be sort of failure and no other.

Acknowledgments I would like to thank John Dupré and Maureen O'Malley for their work on this issue and for establishing and running the Question the Tree of Life research project. I am particularly indebted to my referees and to Maureen O'Malley for helpful suggestions on this paper.

References

- Andam CP, Williams D, Gogarten JP (2010) Natural taxonomy in light of horizontal gene transfer. *Biol Philos.* doi:[10.1007/s10539-010-9212-8](https://doi.org/10.1007/s10539-010-9212-8)
- Baptiste E, Boucher Y (2008) Lateral gene transfer challenges principles of microbial systematics. *Cell* 16:200–207
- Baptiste E, Susko E, Leigh J, MacLeod D, Charlebois RL, Doolittle WF (2005) Do orthologous gene phylogenies really support tree-thinking? *BMC Evol Biol* 5:33. doi:[10.1186/1471-2148-5-33](https://doi.org/10.1186/1471-2148-5-33)
- Baptiste E, Burian RM (2010) On the need for integrative phylogenomics: and some steps toward its creation. *Biol Philos.* doi:[10.1007/s10539-010-9218-2](https://doi.org/10.1007/s10539-010-9218-2)
- Batterman R (2002) *The devil in the details*. Oxford University Press, Oxford
- Beatty J (1995) The evolutionary contingency thesis. In: Wolters G, Lennox J (eds) *Concepts, theories and rationality in the biological sciences*. University of Pittsburgh Press, Pittsburgh
- Beiko RG (2010) Gene sharing and genome evolution: networks in trees and trees in networks. *Biol Philos.* doi:[10.1007/s10539-010-9217-3](https://doi.org/10.1007/s10539-010-9217-3)
- Beldade P, Koops K, Brakefield P (2002) Modularity, individuality, and evo-devo in butterfly wings. *Proc Natl Acad Sci USA* 99:14262–14267
- Bouchard F (2010) Symbiosis, lateral function transfer and the (many) saplings of life. *Biol Philos.* doi:[10.1007/s10539-010-9209-3](https://doi.org/10.1007/s10539-010-9209-3)
- Boucher Y, Baptiste E (2009) Revisiting the concept of lineage in prokaryotes: a phylogenetic perspective. *BioEssays* 31:526–536
- Cairns-Smith AG (1985) *Seven clues to the origin of life: a scientific detective story*. Cambridge University Press, Cambridge
- Ciccarelli F, Doerks T, von Mering C, Creevey C, Snel B, Bork P (2006) Towards automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287
- Coyne J (2009) *Why evolution is true*. Viking, New York
- Dagan T, Martin W (2006) The tree of one percent. *Genome Biol* 7(10):118. doi:[10.1186/gb-2006-7-10-118](https://doi.org/10.1186/gb-2006-7-10-118)
- Dagan T, Martin W (2007) Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc Natl Acad Sci USA* 104:870–875

- Dagan T, Artzy-Randrup Y, Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA* 105(29):10039–10044
- Danchin A (2002) *The Delphic boat*. Harvard University Press, Cambridge
- Darwin C (1859) *The origin of species by means of natural selection*, 1st edn. John Murray, London
- Doolittle WF (1996) At the core of the archae. *Proc Natl Acad Sci USA* 93:8797–8799
- Doolittle WF (1999) Phylogenetic classification and the universal tree. *Science* 284:2124–2128
- Doolittle WF (2004) W. Ford Doolittle. *Curr Biol* 14:R176–R177
- Doolittle WF (2005) If the tree of life fell, would we recognize the sound? In: Sapp J (ed) *Microbial phylogeny and evolution: concepts and controversies*. Oxford University Press, Oxford, pp 119–133
- Doolittle WF (2009a) The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. *Philos Trans R Soc Lond B Biol Sci* 364:2221–2228
- Doolittle WF (2009b) Eradicating typological thinking in prokaryotic systematics and evolution. *Cold Spring Harb Symp Quant Biol* doi:10.1101/sqb.2009.74.002
- Doolittle WF, Baptiste E (2007) Pattern pluralism and the tree of life hypothesis. *Proc Acad Nat Sci USA* 104:2043–2049
- Doolittle WF (2010) The attempt on the life of the tree of life: science, philosophy and politics. *Biol Philos*. doi:10.1007/s10539-010-9210-x
- Forgacs G, Newman S (2005) *Biological physics of the developing embryo*. Cambridge University Press, Cambridge
- Fournier G, Huang J, Gogarten JP (2009) Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. *Philos Trans R Soc Lond B Biol Sci* 364:2229–2239
- Franklin LR (2007) Bacteria, sex and systematics. *Philos Sci* 74:69–95
- Galtier N, Daubin V (2008) Dealing with incongruence in phylogenetic analyses. *Philos Trans R Soc Lond B Biol Sci* 363:4023–4029
- Godfrey-Smith P (2009) *Darwinian populations and natural selection*. Oxford University Press, Oxford
- Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 19:2226–2238
- Gould SJ (1989) *Wonderful life: the Burgess shale and the nature of history*. W. W. Norton & Company, New York
- Gribaldo S, Brochier C (2009) Phylogeny of prokaryotes: does it exist and why should we care? *Res Microbiol* 160:513–521
- Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 54:641–679
- Jain R, Rivera M, Lake J (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* 96:3801–3806
- Johnson K (2007) Natural history as stamp collecting: a brief history. *Arch Nat Hist* 34:244–258
- Kitcher P (1999) The hegemony of molecular biology. *Biol Philos* 14:195–210
- Kitcher P (2001) *Science, truth, and democracy*. Oxford University Press, Oxford
- Labov W (1994) *Principles of linguistic change, vol. 1: internal factors*. Blackwell, Oxford
- Lake J, Moore J, Simonson A, Rivera M (2005) Fulfilling Darwin's dream. In: Sapp J (ed) *Microbial phylogeny and evolution: concepts and controversies*. Oxford University Press, Oxford, pp 184–206
- Lawton G (2009) Uprooting Darwin's tree. *New Sci* 201:36–40
- Lerat E, Daubin V, Moran NA (2003) From gene trees to organismal phylogeny in prokaryotes: the case of the *g*-proteobacteria. *PLoS Biol* 1:101–109
- Lienae EK, DeSalle R (2009) Is the microbial tree of life verificationist? *Cladistics* 25:1–7
- Loewer B (2009) Why is there anything except physics? *Synthese* 170:217–273
- Lopez P, Baptiste P (2009) Molecular phylogeny: reconstructing the forest. *C R Biologist* 332:171–182
- Malkiel Y (1967) Every word has its own history. *Glossa* 1:137–149
- McMahon A (1994) *Understanding language change*. Cambridge University Press, Cambridge
- Michod R (1999) *Darwinian dynamics*. Princeton University Press, Princeton
- Nowak M (2006) *Evolutionary dynamics: exploring the equations of life*. Harvard University Press, Cambridge
- Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304
- O'Malley M, Dupré J (2007) Size doesn't matter: towards a more inclusive philosophy of biology. *Biol Philos* 22:155–191

- Philippe H, Douady C (2003) Horizontal gene transfer and phylogenetics. *Curr Opin Microbiol* 6:498–505
- Phillips B (2006) Word frequency and lexical diffusion. Palgrave Macmillan, England
- Pinker S (1994) The language instinct: how the mind creates language. HarperCollins, New York
- Planet P (2002) Reexamining microbial evolution through the lens of horizontal transfer. In: DeSalle R, Giribet G, Wheeler W (eds) *Molecular systematics and evolution: theory and practice*. Birkhauser, Basel
- Puigbò P, Wolf YI, Koonin E (2009) Search for a 'tree of life' in the thicket of the phylogenetic forest. *J Biol* 8:59. doi:[10.1186/jbiol159](https://doi.org/10.1186/jbiol159)
- Raff R (1996) The shape of life: genes, development, and the evolution of animal form. University of Chicago Press, Chicago
- Shapiro J, Dworkin M (1997) Bacteria as multicellular organisms. Oxford University Press, New York
- Simon H (1996) The sciences of the artificial, 3rd edn. MIT Press, Cambridge
- Sterelny K, Griffiths P (1999) Sex and death. University of Chicago Press, Chicago
- Strevens M (2008) Depth: an account of scientific explanation. Harvard University Press, Cambridge
- Thaler D (2009) The cytoplasmic structure hypothesis for ribosome assembly, vertical inheritance, and phylogeny. *BioEssays* 31:774–783
- Toulmin S (1953) The philosophy of science: an introduction. Hutchinson's University Library, London
- Wilson EO (1994) Naturalist. Grand Central Publishing, New York
- Wilson J (1999) Biological individuality: the identity and persistence of living entities. Cambridge University Press, Cambridge
- Wilson R (2005) Genes and the agents of life: the individual in the fragile sciences. Cambridge University Press, Cambridge
- Woese C (2002) On the evolution of cells. *Proc Natl Acad Sci USA* 99:8742–8747
- Woese C (2005) Evolving biological organization. In: Sapp J (ed) *Microbial phylogeny and evolution: concepts and controversies*. Oxford University Press, Oxford, pp 99–118
- Yates T, Salazar-Bravo J, Dragoo J (2004) The importance of the tree of life to society. In: Cracraft J, Donoghue M (eds) *Assembling the tree of life*. Oxford University Press, Oxford, pp 7–17
- Zhaxybayeva O, Gogarten JP (2004) Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet* 20:182–187
- Zhaxybayeva O, Lapierre P, Gogarten JP (2004) Genome mosaicism and organismal lineages. *Trends Genet* 20:254–260