Forthcoming in *Studies in History and Philosophy of Biological and Biomedical Sciences* (subject to possible editorial revisions).

When Integration Fails: Prokaryote Phylogeny and the Tree of Life

Maureen A. O'Malley
Department of Philosophy, University of Sydney, Quadrangle A14, NSW 2006, Australia
Email: maureen.omalley@sydney.edu.au

Abstract

Much is being written these days about integration, its desirability and even its necessity when complex research problems are to be addressed. Seldom, however, do we hear much about the failure of such efforts. Because integration is an ongoing activity rather than a final achievement, and because today's literature about integration consists mostly of manifesto statements rather than precise descriptions, an examination of unsuccessful integration could be illuminating to understand better how it works. This paper will examine the case of prokaryote phylogeny and its apparent failure to achieve integration within broader tree-of-life accounts of evolutionary history (often called 'universal phylogeny'). Despite the fact that integrated databases exist of molecules pertinent to the phylogenetic reconstruction of all lineages of life, and even though the same methods can be used to construct phylogenies wherever the organisms fall on the tree of life, prokaryote phylogeny remains at best only partly integrated within tree-of-life efforts. I will examine why integration does not occur, compare it with integrative practices in animal and other eukaryote phylogeny, and reflect on whether there might be different expectations of what integration should achieve. Finally, I will draw some general conclusions about integration and its function as a 'meta-heuristic' in the normative commitments guiding scientific practice.

Keywords

Integration, Phylogeny, Tree of life, Philosophy of scientific practice, Heuristic

Highlights

- Integration occurs in regard to data, methods and explanations;
- Prokaryote phylogeny struggles to be integrated into universal phylogeny from all three perspectives;
- In animal and other eukaryote phylogeny, similarly non-integrative situations are found;
- The problems of prokaryote phylogeny can be understood within a meta-heuristic account of integration.

Outline of article

Forthcoming in *Studies in History and Philosophy of Biological and Biomedical Sciences* (subject to possible editorial revisions).

1. Introduction

> '[I]ntegrative biology is both an approach to and an attitude about the practice of science. Integrative approaches seek both diversity and incorporation. They deal with integration across all levels of biological organization, from molecules to the biosphere, and with diversity across taxa, from viruses to plants and animals. Integrative biology provides both a philosophy and a mechanism for facilitating science at the interfaces of "horizontally" arrayed disciplines, in both research and training' (Wake, 2003, p. 240)
>
> 'Think of integrative biology as biology from a big picture point of view in which the relationships between parts are studied in order better to understand the whole. From genomics to global change, integrative biology seeks to discover the complex interrelationships between living organisms and the physical and biological environment in which they live. This is the new biology, with an emphasis on bringing multiple disciplines to bear on complex scientific questions' (School of Integrative Biology at the University of Illinois at Urbana-Champaign, http://sib.illinois.edu/WhatIsIB.htm)

These quotes and many more just like them show that integration is an important concept in today's life sciences. Not just important, but almost self-evidently necessary: who would deny the desirability of a more multidimensional picture of biology? For the more cynical reader, however, integration could well be just another one of those buzzwords in biology, similar to 'interdisciplinary' or 'non-reductionist', with vague referents and a function simply to encourage a valued outcome rather than describe its constituent activities. In what follows, I will suggest that integration is a specifiable set of activities that enables researchers to analyse diverse and extensive datasets in relation to multilevel research questions. In previous work, I have identified integrative practices from a positive point of view: the things scientists do to achieve integration (O'Malley & Soyer, 2012). In this paper, I will approach integration from a negative point of view to try and sort out what is going on when integration appears not to work. I will use the example of prokaryote phylogeny because of the obstacles this field has faced in its aim to be integrated within 'universal' phylogeny. At the end of this discussion, I will put the positive and negative aspects back together again to see if this produces a more nuanced account of integration and what its implications are for normative accounts of scientific practice.

2. Integration in the life sciences and philosophy

Philosophers of science aim, broadly speaking, to understand how science works. Recently, this aim has been qualified as a shift from thinking of knowledge as a finished theoretical product to knowledge generation as an ongoing and diverse set of social epistemic practices (Brigandt, forthcoming). Many earlier philosophical efforts focused on 'unification' as a major contributor to the dynamic process of science. Scientific unification (understood here as a subject distinct from the metaphysics of unity/disunity) has been conceived methodologically and epistemologically, with far greater attention given to the latter (Wylie, 1999; Morrison, 2000). Epistemologies of unification have until recently focused positively and negatively on a single posited mechanism of that unification – the reduction of one theory to another (Oppenheim & Putnam, 1958; Kitcher, 1981; Schaffner, 1993). These days, however, philosophers of science have begun to use the term 'integration' to describe aspects of scientific practice that had previously been collapsed into the more abstract notion of unification. This new work suggests that integration is central to an understanding of how fields and disciplines, bodies of data, combinations of methods, and different levels of explanation work together to expand knowledge and produce innovations in scientific practice (Brigandt & Love, 2012; Brigandt, 2010; Leonelli, 2008; Grantham, 2004a; Mitchell, 2003; Morrison, 2000; Wylie, 1999; Bechtel, 1993). While often these inquiries are focused on specific activities, there is also a more general notion of integration at play in philosophy and scientific practice. Philosopher Todd Grantham (2004a; this issue; see also Kitcher, 1999) discusses this broader notion of integration as a 'regulative ideal'. I will suggest later in this paper that it can also be understood as a 'meta-heuristic' or a guiding strategy that provides insight into the research phenomena even when integrative practices appear to fail.

In most life science research, the term 'integration' refers to specific activities by which diverse methods, bodies of data and models are brought together in order to gain a mechanistic and predictive understanding of biological systems (Liu, 2005). Integration is a focus of attention in the molecular life sciences, particularly when such research attempts to combine new molecule-based knowledge with existing knowledge derived from morphological and physiological data. Molecular biology has seen a proliferation of quantifiable molecular data that have yielded limited insight to single-disciplinary approaches. Individual efforts have given way to integrated projects carried out by multidisciplinary teams of molecular, evolutionary and cell biologists, biochemists, bioinformaticians and other computational scientists, mathematicians, engineers and physicists (Lauffenburger, 2012; Aderem, 2005; Ripoll et al., 1998). Departments, institutes, societies and journals have been acquiring labels of 'integrative biology' for over a decade now, thus meeting many of the social criteria for disciplinary status (Powell et al., 2007; Gerson, this issue).

But despite the apparent scientific necessity and social value of integration, scientific literature tends to address integration very broadly, sometimes not specifying what it entails (e.g., Wake, 2008). While a single precise definition of integration is not by any means the aim of this paper, it is important to identify what is meant by the term and whether different interpretations can be combined coherently into a general use of the concept. In previous work on integration in molecular systems biology, Orkun Soyer and I suggested that integration occurs in three distinct but interconnected ways (O'Malley & Soyer, 2012). *Data integration* is the response to the masses of molecular and other biological data often collected without specific hypotheses in mind. Combining large datasets involves database modelling, accurate data quantification, standardization procedures, and the design of user interfaces that enable data to be combined in novel ways and reanalysed for new research questions (Lenzerini, 2002; Cali et al., 2004; Ideker et al., 2007; Ghosh et al., 2011). *Methodological integration* involves directing a range of methods at a particular biological phenomenon or research problem in order to achieve multiple perspectives on how a system works or what the dimensions of the problem are. It is presumed that these combinations of methods and methodologies can produce knowledge that is not obtainable from single-method or even single-discipline approaches (Mykles et al., 2010; Hyman 2011). *Explanatory integration* refers to the synthesis of previously unconnected theories and the import of explanatory and predictive models from other research domains into new areas of inquiry (Patel & Nagi, 2010). These models can be mathematical or statistical representations of biological systems, as well as the traditional conceptual models with which molecular biologists are more familiar (Brigandt, this issue). The notion of explanatory integration does not, however, involve the goal of a complete, unified explanation of all biology, which is what philosophers have usually had in mind when discussing 'theoretical unification'. Integration does, however, have a broader normative function, which encourages connections between related bodies of research to achieve epistemic and practical advantages (Grantham, 2004a; this issue; Burian, 1993).

Although there are many success stories of integration (due to integrative practices transforming understandings of particular biological systems – see O'Malley & Soyer, 2012), these successes do not mean that the problems of achieving integration are not well recognized in the scientific literature. There is considerable concern, for example, about how to implement integration in regard to data. Experimental techniques and procedures can produce highly variable results that are hard to model in any generalizable way, and combining different data types remains a stubbornly persistent problem (Schilling et al., 2008; Sullivan et al., 2010). Methods that produce insight into some phenomena and research questions may have to be applied very cautiously to different systems and for different purposes (Hyman, 2011). Likewise, integrating theoretical resources from one domain to another can produce misleading and even artefactual models (Prill et al., 2010). Another potential challenge for integration comes from disciplinary relationships and boundaries. While some biologists do talk about 'disciplinary integration' (e.g., Roth, 1994; Auffray et al., 2003; Chuang et al. 2010), disciplines are more likely to form the contexts in which methods, data and explanations are integrated. As I have argued previously, multidisciplinary capacities certainly contribute to and even guide integration, but they function as conditions for integration rather than integration itself (O'Malley & Soyer, 2012; Bechtel, 1993). In other words, disciplines do not go away in integrative research. They may become even more important because of the specific and highly specialized contributions that individuals need to make to an integrative research project. Nevertheless, disciplinary context can greatly affect integrative practices in regard to data, method and model integration, and this context needs to be taken into account in order to understand how integration actually works.

Because integration is deemed so important in many life sciences, closer scrutiny of the activities and conditions necessary for its achievement would be useful. Does integration involve, for example, 'more' of everything, as the quotes at the beginning of the introduction (Section 1) might imply? All discussions of integration, whether philosophical or scientific, presume integration is necessary and should be increased, and that to be successful, integration simply needs appropriate motivation and the right tools. However, criteria of successful integration are seldom specified even by its most ardent advocates. The mechanisms of integration, whether in regard to data, methodological or explanatory integration, are not known in a detailed yet generalizable way, and broad claims about 'disciplinary integration' and its importance do not clarify how integration works. Paying attention only to fields such as molecular systems biology, where integration is assumed to be both necessary and successful (although no fine-grained study actually demonstrates this), is likely to produce self-confirming bias. One way in which such bias could be avoided and a more thoroughgoing account of integration produced is by focusing on fields where integration fails or at least appears to fail.

3. Phylogeny and the universal tree of life: the prokaryote problem

One of the major achievements of the massive 'omic' datasets generated in the last two decades is the capacity for comparative analysis they have bestowed on the biological sciences. Molecules can be compared not just from organism to organism (e.g., gene expression in genomically identical cells), but across a huge range of organismal lineages. Molecular phylogeny, which gained its early successes on the basis of hard-won single-molecule comparisons in the 1960s (Pauling & Zuckerkandl, 1963; Dayhoff et al., 1974), has transformed phylogeny in all of its subfields. This was especially the case for evolutionary microbiology. Microbiologists had long felt deprived of sufficient morphological characters, confounded by the biochemical flexibility of microorganisms, and bedeviled by obdurate difficulties in achieving any semblance of 'natural' species classifications (Stanier and van Niel, 1962). They enthusiastically embraced molecular sequences as a means by which to bring microbes and particularly prokaryotes into a Darwinian system of classification (Woese, 1987). Once universal characters of nucleotides and amino acids had become the basic data for categorizing microorganisms, the Darwinian vision of a universal representation of the evolutionary history of all life – the Great Tree of Life (Darwin 1872, p. 105) – shifted from a disciplinary 'dream' to an imminent reality (Woese, 1996).

The tree of life is usually conceived as a composite representation of all evolutionary lineages of organisms (not viruses or plasmids or other evolving genetic elements), extinct and extant. This tree depicts how all life is related and where every divergence between lineages has occurred. Such a tree is necessarily a *universal* phylogeny because it has to incorporate all evolved and evolving life. Branching patterns are deemed to capture speciation, and the branches themselves represent species. As well as enabling classification, the tree of life explains in a 'natural' (i.e., not purely pragmatic) way why such classification obtains: because of evolutionary processes over the depth of evolutionary time. Obviously, any such representation is currently incomplete because many lineages have yet to be mapped to the tree, and many relationships resolved. However, the tree of life exists both as an abstraction of evolutionary reality and an epistemic aim that lies at the heart of the evolutionary biology community. As Richard Dawkins so bluntly exclaimed, 'everyone' knows this tree actually exists, and he and many others expect eventually to be able to depict it in its entirety (Dawkins, 2003, p. 112). That was also the mood and motivation in molecular microbial phylogeny in the 1980s as leaders in molecular methods proclaimed that molecules would soon allow the realization of a Darwinian representation of all evolutionary relationships (Woese, 1987).

But more or less immediately upon the turn to molecular analysis, this new phylogenetic approach generated controversy. Suddenly, on the basis of molecular data, the biological world was no longer divided into five 'obvious' kingdoms of plants, animals, fungi, protists and monera (Whittaker, 1969), or even more fundamentally, into the two lifeforms of prokaryote and eukaryote (Stanier & van Niel, 1962; Mayr 1982). Instead, molecular analyses carved the biological world into three groups that reflected the major evolutionary trajectories of life: Archaea, Bacteria and Eukarya (Woese & Fox, 1977; Woese, 2005). With this 'primary tripartite division of the living world' (Woese et al., 1990, p. 4576), molecular phylogeny had achieved a single global representation: the 'universal phylogeny' into which all new and existing knowledge about evolutionary process and pattern could be integrated (Woese, 1987; 2000; Sidow & Wilson, 1990; Wilson et al., 1977). The tree of life could thus function as both an evolutionary classification of organisms and an evolutionary explanation of those relationships.

Alas, these declarations and general rejoicing rapidly began to seem premature. Even before the availability of large molecular datasets, the phenomenon of lateral gene transfer (LGT, synonymous with horizontal gene transfer) began to pose problems for a *unique* universal phylogeny with its pattern of ever-bifurcating branches. Initially explained as a process that occurred due to laboratory manipulations, microbial phylogeneticists and other biologists quickly realized that LGT played an important biological and evolutionary role in a range of organismal lineages (Anderson, 1968; Coughter & Stewart, 1989; Jones & Sneath, 1970; Reanney, 1977). But even when this was known, gene transfers continued to be minimized conceptually and methodologically as 'anomalies' that did not fundamentally endanger the basic tree structure (e.g., Wilson et al., 1977; Schwartz & Dayhoff, 1978; Wheelis et al., 1992; Woese et al., 1990). It did not take long, however, for LGT to be conceptualized in a more radical way, as a foundational problem

Forthcoming in *Studies in History and Philosophy of Biological and Biomedical Sciences* (subject to possible editorial revisions).

for the total tree of evolutionary history, due to the web-like connections it made between branches (Hilario & Gogarten, 1993; Figure 1).

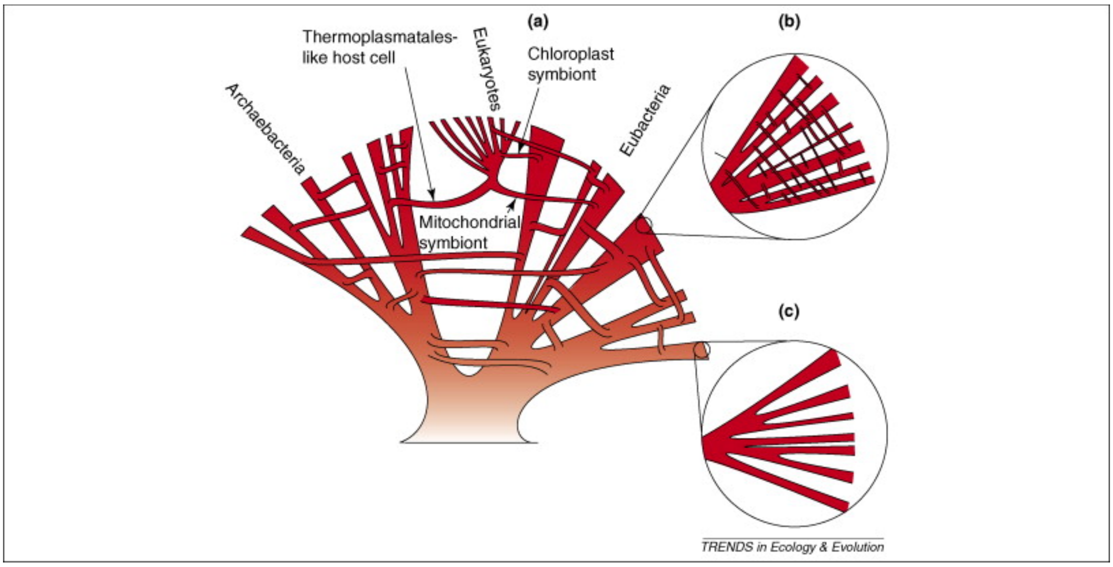

Figure One
A network of life: a) Chimeric eukaryotes, with only the lateral integrations of whole cells displayed (see Section 3 for discussion); b) LGT between bacteria, occurring to the extent that these groups are effectively panmictic (recombining unrestrictedly); c) absence of detectable LGT in some bacterial groups. Overall, vertical descent is still emphasized and thus the tree topology (not the case for all network representations). Used with permission from MacInerney et al., 2008 (figure legend paraphrased).

The more indispensible molecular analysis became, the more findings were made that LGT was both rampant (common) and promiscuous (without regard for the fidelity of species or even broader classification boundaries). Different gene trees were increasingly found to be incongruent with trees made of sequences from the same organisms; gene trees contradicted basic biological rationales behind standard organism-based classifications (Doolittle, 1999; Martin, 1996; 1999). Even conservative phylogeneticists, committed to the notion of universal phylogeny, made and reported such findings (e.g., Woese, 2000). The plethora of mobile genetic elements participating in these transfers (plasmids, transposons, phage genomes, and other modules) were eventually given their own database and classification system (Leplae et al., 2004).

In a further twist of epistemic fate, the whole-genome data that enabled even broader comparative evolutionary analyses – often called phylogenomics, and expected to overcome once and for all problems of incongruence – exposed even more evidence of the mosaicism of prokaryote genomes (Gogarten et al., 2002; O'Malley & Boucher, 2005). Acquired DNA from other lineages was found to comprise up to 35% of some organism's genomes (e.g., Lawrence & Ochman, 1998; Nelson et al., 1999; Deppenmeier et al., 2002). Even in the very same taxon (i.e., a 'species'), different strains ('sub-species') were discovered to vary hugely in gene content and to exhibit extremes of phenotypic diversity, e.g., virulence and avirulence (Welch et al., 2002; Medini et al., 2005; Tettelin et al., 2008; Lapierre & Gogarten, 2008; Lukjancenko et al., 2010). Within populations, intraspecies homologous recombination – the process by which microrganisms share small amounts of similar but not always identical DNA – added further complications: clonal reproduction and genetic identity within populations could no longer be assumed (Maynard Smith et al. 2000; Feil et al. 2001; Lawrence, 2002). Different taxa have very different recombination rates, even within the same supposed species, making general calculations of this phenomenon difficult (Didelot & Maiden, 2010; Boucher & Bapteste, 2009).

Consequently, when the evolutionary history of an organismal lineage is inferred from molecular data, only tiny amounts of some genomes show indications they have *not* been transferred over the evolutionary long run. Those minuscule untransferred 'cores' either do not overlap with other organisms or reveal far too little of a shared evolutionary history to say anything general (Dagan & Martin, 2006; Bapteste et al., 2008). Strong habits of gene exchange may in fact create what are perceived as patterns of vertical descent because of preferential gene transfer between more closely related organisms (Andam et al., 2010). Instead of more sequence data adding resolution to blurred or conflicting branching patterns, such data can obscure the tree of life and make the state of phylogenetic knowledge (in the form of uniquely bifurcating branches) arguably worse than before the molecular era. Although from a broader evolutionary view, knowledge of lateral exchanges of genetic and cellular resources (such as the mitochondrion in Figure 1) adds new dimensions to the evolutionary picture, these phenomena do not fit at all the expected pattern of ever-diverging branches.

Nevertheless, the insights molecules have afforded evolutionary analysis, whether microbial or macrobial, make these data, methods and associated concepts too valuable even to consider relinquishing. Instead, a range of strategies has been developed to dispose of LGT problems. Amongst them are: removing LGT-prone genes and obvious mobile genetic elements, such as plasmids and viruses, from the analysis; conceptually discounting the evolutionary importance of such transfers so they can be legitimately ignored; and designing an assortment of methods that can focus exclusively on vertical signal (Galtier & Daubin, 2008; Bapteste et al., 2009). In addition, universal phylogeny has many other problems besides (or accompanying) those of LGT. A major one is the extraction of patterns of vertical descent from all other molecular signals. Doing this requires methods that can separate and prioritize tree-like patterns from non-tree-like patterns, as well as from poor signal and phylogenetic artifacts. Even if this is done rigorously, the resultant tree may be neither the history of any single gene nor the history of the organism itself because of the way 'average' signals are constructed and the selection of evolutionary information this requires (Haggerty et al., 2009; Bapteste et al., 2009; Ragan & Beiko, 2009). Since 'highways of gene sharing' are deeply informative of historical ecologies and selection pressures (Beiko et al., 2005; Zhaxybayeva et al., 2009), this process of data selection means that excluding these horizontal movements removes major evolutionary insights.

Although evolving prokaryote lineages behave in ways that are obviously difficult to accommodate within a universal phylogeny, eukaryotes have also been evolutionarily wayward. The endosymbiosis that is a primary characteristic of all eukaryote cells (the incorporation of another bacterium in the eventual form of the mitochondrion) is a massive horizontal event occurring at the very base of the eukaryote tree (Archibald, 2011; see Figure 1). It involved first the engulfment of a cell, and then the gradual import of many of the engulfed cell's genes into the host's nucleus – a process known as endosymbiotic gene transfer or EGT (Martin & Herrmann, 1998). Other major endosymbioses have occurred several times in the evolutionary history of eukaryotes (primary, secondary and tertiary plastid endosymbioses), meaning that the path of eukaryote-hood is fundamentally marked by lateral, non-tree-like processes (Martin 2011; Archibald 2012; Moustafa et al., 2009; Pisani et al., 2007). Hybridization, another reticulating event, is common in many multicellular eukaryotes and has been an instigator of the speciation that is caused by lineages merging rather than splitting (Mallet, 2005; Arnold, 2007). Moreover, many prokaryote to eukaryote transfers have been detected (Andersson, 2009; Hotopp et al., 2007; Loftus et al., 2005), and even some from eukaryotes to prokaryotes (Keeling & Palmer, 2008). Gene exchanges have occurred between eukaryotes too, and the more eukaryote genomes are analysed the more such exchanges are found (Andersson, 2009; Richards et al., 2009). But the relative infrequency of LGT and hybrid speciation in the eukaryote branches of the tree of life means for most evolutionary biologists that these reticulations can still be justified as minor phenomena when compared to vertical descent in these taxa (Keeling, 2009; Bapteste et al. 2009; Dagan & Martin, 2009).

One solution to the LGT problem might be to think that prokaryote evolution should not be integrated into the 'general' study of evolutionary process and pattern. This exclusion could be justified by an argument that data and methodological integration might have occurred in prokaryote phylogeny but not explanatory integration, and that explanatory adjustments of some sort should be sought so that prokaryote evolution is explained separately (i.e., non-integratively). The tree of life could, for example, be conceived as the representation of a major trend explained by the evolutionary processes experienced by most multicellular eukaryotes (O'Malley, 2010a). This model of vertical descent with endogenous modification would not assimilate or be expected to assimilate the evolution of organisms prone to gene exchange across lineages ('exogenous' modification). Following this reasoning, there should thus be little difficulty in reconstructing the tree of life (recall that it is necessarily a unique tree of ever-bifurcating branches) as a representation of a restricted set of life forms, a limited period of evolutionary history, and a limited number of processes and patterns. Processes left out would include eukaryogenesis (the evolutionary event that merged a bacterium and an archaeon into a completely new type of cell), other endosymbiotic mergers in eukaryotes (including those to do with the chloroplast and other plastids), a variety of eukaryotic gene transfers, much prokaryote evolution, and of course early life (sometimes described as a period of 'unrestrained' transfer). Consequently, this solution would severely delimit 'universal' phylogeny, and could also perturb those for whom integration means more scope rather than less. What the prokaryote problem in universal phylogeny therefore highlights is the means by which consistency and unity are achieved – in this case *not* by integrating in the sense of accumulating, but by excluding certain data, methods and evolutionary explanations.

4. How integration into universal phylogeny fails for prokaryote phylogeny: a comparison with animal (and other eukaryote) phylogeny

This sketch of the conundrum of prokaryote phylogeny vis-à-vis universal phylogeny needs to be broken down further to get a sense of exactly how prokaryote phylogeny fails to be integrated into the bigger evolutionary picture. Distinguishing the three main modes of integration identified above (data, methods, explanation) as well as showing their connections, I will examine a range of integrative efforts being made to circumvent the LGT problem and synthesize the representation of prokaryote and eukaryote evolution. But as already noted, phylogeny in eukaryotes – even animals – is not so straightforward either, and comparing the integrative problems of the latter with those of prokaryote phylogeny will be instructive. Very commonly in eukaryote phylogeny, especially eukaryotes for which there is abundant morphological data (anatomical, developmental, fossil), there can be revealing tensions between phylogenetic narratives inferred from different data types (Grantham, 2004b). What happens when such conflicts arise? How are they resolved? Is there ever any conclusion made that such conflicts might be fundamentally irresolvable, in virtue of the fact that some data tell different evolutionary stories? Answers to these questions could have bearing on how integration has and has not worked for prokaryote phylogeny in relation to universal phylogeny.

4.1. Data integration: inclusion and exclusion

While molecular phylogeny in prokaryotes and eukaryotes relies on similar types of data (e.g., ribosomal molecules for reference trees; genome-wide data for extensive comparison), the data that are not integrated are those from which LGT can be inferred as well as a great deal of other 'phylogeny unfriendly' data, often described as 'noise' or 'non-signal'. Data integration in prokaryote phylogeny, when it is envisaged as a contribution to universal phylogeny, thus involves a commitment to a single evolutionary pattern that will *not* be inferred from the bulk of the molecular data available. What must be identified are genes that have detectable tendencies to be exchanged. Sometimes this is done by putting aside whole classes of problem genes, and other times by ruthlessly removing any single gene that does not fit the expected ‘species’ tree (an assumed pattern of bifurcating relationships indicated by other data, which are usually molecular for prokaryotes).

The class method of exclusion often focuses on distinctions between informational (e.g., transcriptional and translational) and household (e.g., metabolic) functions of DNA. The latter are usually considered to be more common cross-lineage currency in evolution than the former. However, even core informational genes (i.e., ribosomal) are known to have been transferred at least occasionally, sometimes with major phylogenetic effects (Xie et al., 2008; Brochier et al., 2000; Yap et al., 1999; Boussau et al., 2008). And new findings indicate that biological function is not the determining factor behind LGT-proneness. Instead, ‘connectivity’ in the form of protein-protein interactions of the gene products is what limits the flow of genes to other organisms: the more connected, the less transferrable – at least in the standardly fragmented way in which transfers usually occur (Cohen et al., 2011; Gophna & Ofran, 2011). Although genes that show traces of or proclivities for transfer are routinely excluded from phylogenetic analyses, for some commentators this is a presuppositional imposition on the data (Doolittle & Bapteste, 2007; Martin, 2011). It can also be methodologically tricky because of how different methods pick out different sets of discordant sequence (Ragan et al., 2006), and because transferred DNA can actually provide support for expected vertical patterns (e.g., Huang and Gogarten, 2006).

One way to look more closely at the legitimacy and viability of excluding LGT-prone genes is to look at what is going on in phylogenies concerned with the paradigm organisms of evolution, animals. Molecular data have been used to revise the animal tree of life in radical ways. New inferences of relationships have swept away older conceptual assumptions about simple to complex trends (with the elevation to later evolutionary history of groups once thought to be at the base of the animal tree), the naturalness of certain ranks (particularly phyla), and the monophyly of major morphologically defined ‘clades’ (Adoutte et al., 2000; Halanych, 2004; Jenner, 2004b). The first animal-wide molecular phylogenies in the late 1980s used a single ribosomal gene, just as prokaryote phylogeny did – in fact, the landmark paper of the ‘new animal phylogeny’ is co-authored by a number of microbiologists (Field et al., 1988). Animal gene trees, many of which were not congruent if a diversity of genes were used, were superseded by genome-based trees in the phylogenomic era, and this expansion brought about even more dramatic changes to depictions of evolutionary relationships in the animal tree of life. Genome-scale data have indicated to some animal phylogeneticists ‘the end of incongruence’, because each branch of the tree should have ‘unequivocal support from all the data’ (Gee, 2003, p. 782). But just as in prokaryote phylogeny, these large-scale datasets have generated further conflicts between branching patterns, even at the relatively coarse level of phyla, and posed questions about the very resolvability of the animal tree, especially at its base (Bourlat et al., 2008; Telford, 2008). However, in animal and other eukaryote phylogeny, conflicts arise not just between sequence-based trees but also between trees based on different data sources (Griesemer, this issue). In animal phylogeny, the traditional source of data has been morphological and its relationship to molecular data is far from an additive one.

In early molecular animal systematics, congruence between molecular and morphological phylogenies was taken largely for granted because the basic scaffold of the tree had already been given by older morphological analyses. Molecules were simply expected for the most part to fill in and resolve that tree (Edgecombe et al., 2011; Edgecombe, 2010; Caterino et al., 2000). More recently, however, the data has been integrated the other way around: by building a molecular tree and then adorning its tips with morphological characters only afterwards. Simultaneous approaches, better known as total evidence (see below), combine both types of data before analysis and consider this sort of integration superior (Eernisse and Kluge, 1993; Hermsen & Hendricks, 2008; Assis, 2009; Wiens et al., 2010). While there is general agreement that molecular data can clarify relationships previously understood only poorly on a morphological basis, and that morphological data can add to the resolution of molecular phylogenies of animals and other eukaryotes (Wortley & Scotland, 2006; Giribet, 2010: Wiens et al., 2010; Huang et al., 2011), it is also undeniable that the use of purely molecular data is very much in the ascendant.[1]

This ‘hegemony’ is often lamented and just as frequently justified. One epistemological rationale for prioritizing molecular data is that they are ‘objective’, whereas morphological data are ‘subjective’ (Halanych, 2004; see Suárez-Diáz & Anaya-Muñoz, 2008 for further discussion). A well known critique (by plant phylogeneticists but about phylogeny in general) of the preferred use of morphological characters argued that such characters are mostly ‘ambiguous’, and unambiguous ones are few and far between (Scotland et al., 2003). Integrating these relatively few unambiguous morphological characters with the abundance of DNA sequence data is the only way forward according to this argument. ‘Integrated studies’ will thus consist of ‘a few morphological characters in the context of a molecular phylogeny’ (Scotland et al., 2003, p. 54).

[1] A recent meta-analysis of insect systematics has examined which types of data are used to generate phylogenies of arthropods compared to phylogenies of other animals (primarily vertebrates) and plants (Bybee et al., 2009). The authors compared molecular, morphological and combined phylogenies published in major journals between 1992 and 2007 (a total of 1469 phylogenies). 73% of these phylogenies were based on exclusively molecular data (which had become increasingly prominent over the 15 years in all the journals analysed), 18% used only morphological data, and the small remainder (9%) used a combination of data types. These percentages varied slightly according to the groups of organisms the phylogenies were about, but not in any particularly revealing way (except to confirm the well known availability of morphological characters for insects). The authors rejoiced in the fact that morphological data were still used despite molecular data being ‘easier, faster and more cost-effective’ (Bybee et al., 2009, p. 5), but did not identify any consequences of non-combined phylogenies.

But as numerous commentators have noted, molecular data itself abounds with ambiguities, not least in regard to single-gene markers that can, for example, grossly misrepresent important evolutionary relationships such as those between birds, mammals and crocodiles (Jenner, 2004a). We have already seen in prokaryote phylogeny (Section 3) how the potential for molecular markers to mislead is well known. And the original molecular clock hypothesis, by which the calculation of a constant rate of nucleotide substitutions enables the dating of divergence between lineages, has had to be extensively remodelled to accommodate highly variable substitution rates in different lineages, including animal lineages (Welch & Bromham, 2006; Donoghue & Benton, 2007). Models of rate heterogeneity and methods to deal with it were in fact developed because of massive inconsistencies between molecular and morphological date estimates. Having seen rapid progress under the molecular regime, however, many animal phylogeneticists believe that 'major phylogenetic problems … will eventually yield under the weight of more molecular sequence data' (Regier et al., 2008, p. 920; Gee, 2003). But the solution is unlikely to be as simple as piling up more molecular data and presuming the true tree will emerge because even when molecular phylogeneticists use the largest possible datasets, incongruence persists (Philippe et al., 2011; Rokas & Carroll, 2006). Evolutionary relationships between many animal groups of different ranks are still unresolved or poorly supported, whether molecular, morphological or combined data sets are used (Jenner, 2011). Even though the animal tree of life is now widely accepted to consist of five major clades, the relationships between these groups are still unclear (Edgecombe et al., 2011; Dunn et al., 2008).

Given the difficulties of constructing a fully resolved animal tree, a common suggestion is that data that are phylogenetically 'unreliable' (showing non-vertical processes) can justifiably be eliminated, and taxa that behave 'badly' in evolutionary analyses should be 'culled' or at least treated specially (Jeffroy et al., 2006; Paps et al., 2009). This justification is made even under the rubric of the 'total evidence' approach, which is interpreted to mean 'all relevant evidence' conjoined with the removal of 'misleading data' (Lecointre & Deleporte, 2005). According to this approach, all 'potentially informative characters' (Eernisse and Kluge, 1993) should be combined, thus uniting molecular and morphological, fossil and living organism data. Total evidence is justified on the grounds that it is the least assumption-laden approach, and a maximally descriptive and explanatory one (Eernisse and Kluge, 1993). Lateral transfers in whichever organisms they occur are theorized as misleading and minor phenomena in light of the aim of tree construction (Lecointre & Deleporte, 2005). However, as has been pointed out for many phylogenetic efforts, the integration of stratigraphic or fossil data with living organism data (molecules and morphology) has often failed because of the lack of methods able to evaluate qualitatively distinct data types against different phylogenetic hypotheses (Grantham, 2004b; Lockhart & Cameron, 2001). Likewise, in prokaryote phylogeny, tree building with ever-larger datasets of genes does not yet have methods available to deal with the scale of the data and amounts of incongruence (Leigh et al., 2011).

The persistence of such issues even in eukaryote phylogeny shows that integrating data is not a straightforward matter of the more the better, whether it is more of the same data, or more types of data. Not only do methods need to be developed to combine different data types, but also epistemic rationales need to be given for identifying certain data as irrelevant or uncombinable. Selecting epistemically appropriate data is definitely a pragmatic necessity in all scientific fields, but in the case of phylogeny the overarching presumption that a tree and only a tree should be the result of the analysis clearly restricts the way data integration works for prokaryotes and other organisms. Whatever the prognosis for animal phylogeny, it has the same underlying issues as in prokaryote phylogeny (and thus universal phylogeny), whereby identifying a particular trend in the combined evidence is enough to justify the tree it is made to produce.

For all phylogeny, however, clashes between and amongst molecular and morphological data can lead to the development of more innovative methods, novel hypotheses about evolutionary relationships (i.e., in order to explain the unexpected clashes), and even the creation of new categories of data. For example, conflicts between morphological and molecular trees in families of corals have resulted in the division of morphological data into macro-morphological and micro-morphological, with the latter generating trees that are by and large congruent with molecular trees (Budd & Stolarski, 2011). Despite success stories such as these in which data are recategorized and made to produce congruent trees, and notwithstanding a general optimism about reconstructing the definitive tree of animals and other eukaryotes (e.g., Telford & Copley, 2011), some eukaryote phylogeneticists doubt that even the availability of 'total' data (referring here to whole genomes of every taxon) will produce 'the full branch structure of the tree of life' (Scotland et al., 2003, p. 543). 'Given that we can never gain access to the "one true tree of life", by definition we cannot assess its accuracy, when it is an absolute rather than a relative property' (Bateman et al., 2006, p. 3411). Such uncertainty and pessimism, while unlikely to be characteristic of the majority of eukaryote phylogeneticists, certainly resonates with similar minority sentiments in prokaryote phylogeny.

Prokaryote phylogeneticists have very little morphological data to combine with molecular data, and thus have even stronger expectations that genomic databases are key to the end of incongruence (e.g., Klenk and Göker, 2010). While data integration is as much about the choice of which data to use as it is about combining data, choices are methodologically and conceptually constrained in different ways. Animal phylogeny is able to make some major evolutionary assumptions because of the biology of these organisms and the evolutionary traces afforded by this biology (e.g., fossils, anatomy), whereas prokaryote phylogeny is largely restricted to combinations of different molecular datasets. Nevertheless, the same process of methodological and conceptual selection of data applies to all forms of phylogeny, and the question becomes one of how such selections are justified. For those who choose to preserve the tree of life, evidence is restricted to congruent tree-producing data; for those who are more concerned to interpret the plenitude of available data without tree-building constraints, a different evolutionary representation is inevitable. Data integration is difficult in eukaryote phylogeny, despite well accepted explanations of expected tree patterns; in prokaryote phylogeny, data integration is even more problematic because of the underlying disagreements over appropriate methods and explanations.

4.2. Methodological integration: multiple methods, different aims

While it is usually taken for granted that traditional phylogeny should employ tree-building methods of inference and representation (because of the fact bifurcating lineages are the only pattern recognized in post-cladist phylogeny; reticulation events including hybridization are not considered to be phylogenetic patterns and nor are multifurcations), it is increasingly recognized in prokaryote phylogeny that tree methods need to be supplemented or even replaced by network-building methods (Lapointe et al., 2010). All trees are networks, but not all networks are trees; networks are therefore more informative of a plurality of evolutionary processes than trees. Bifurcating branches are just one pattern that can be isolated from a wider set of relationships. Networks effectively include reticulation events and multiple furcations from the same node, while still encompassing the bifurcating patterns explained by vertical descent. Since reticulate evolution cannot be forced into tree topologies, methods that can represent networks as well as trees have used this as the warrant for their development (Huson & Bryant, 2006; Baroni et al., 2006; Dagan et al., 2008). Network methods can be used for genes or whole genomes, and represent conflicts between data – not all of which are due to historical reticulation events (Beiko, 2010; Huson & Scornavacca, 2010; Morrison, 2010). Networks can strategically enable advances in phylogenetic efforts by delimiting the possible evolutionary trajectories between lineages (Beiko, 2010). But despite the proliferation of network constructing tools, the most common methods by far are still tree-based even in prokaryote phylogeny, at least in part because network methods still require considerable development (Morrison, 2010; Woolley et al., 2008; MacInerney et al., 2011). Over-connected networks (‘hairballs’), which are often produced when large amounts of genome data are analysed, are very hard to visualize and interpret effectively (Beiko, 2011).

For some plant phylogeneticists, the necessity of integrating network and tree methods is clear: without including hybrid speciation, evolutionary relationships amongst plants will be incompletely and inaccurately understood (Linder & Rieseberg, 2004; Vriesendorp & Bakker, 2005). Networks can be used in animal phylogenetics too, where they can indicate much more inclusively than standard tree methods both the evolutionary history of the lineages involved and where there is data discordance (Huson & Bryant, 2006). But methodological integration in animal phylogeny is still focused firmly on trees, with a more common integrative aim being the production of trees from multiple rather than single tree-building methods (Jenner, 2011). A large part of the rationale for combining methods is that they will settle on a reconstruction that is closer to the actual (assumed) tree of life than can be achieved by any individual method.

A good example of this type of synthesis is found in fungi phylogeny, which has been as disadvantaged as prokaryotes and much of the plant kingdom in its shortage of fossil data. Recent efforts to produce an adequate tree of fungi not only used two different sets of genome-wide data but four tree construction methods in order to achieve consistent interpretations (Ebersberger et al., 2011). The authors saw the resultant tree as a sketch of the fungal tree of life that could be refined with more data, rather than as a definitive statement of evolutionary relationships. Given the evolutionary and phylogenetic complications of fungi (LGT, plus lateral chromosome transfer, hyphal fusion with hundreds of different nuclei in one cell, and large numbers of fungi forming intimate symbioses with algae as lichens) this admission of the fungal tree as not 'definitive' can be read as rather an understatement from a broader evolutionary perspective. But by using multiple methods to home in on the 'backbone' tree, evolutionary mycologists and phylogeneticists more generally can take heart that progress has been made in the fungal part of the universal phylogeny.

The same can be said of animal phylogeny, by contrasting early animal-wide molecular trees with trees produced from phylogenomic data. It is not only the data that have ratcheted up in quantity but also the combinations of methods (DeSalle & Schierwater, 2008). The aim of multi-method combinations is to discern ever more clearly the 'true' phylogenetic signal and neither incorporate nor artefactually create non-phylogenetic signal (Philippe et al., 2011). However, in prokaryote phylogeny multiple methods are sometimes used not in a synthesizing way but as alternatives, by making use of appropriate data to 'reveal a multitude of alternative affinities' between lineages (Beiko, 2011). In this sort of methodological integration, the synthesis of different findings into one common representation is *not* presumed. This pluralistic interpretation is at odds with the notion of universal phylogeny, which traditionally assumes a single true tree of unique branching patterns as discussed above (Section 3).

A major factor guiding method choice and integration in phylogeny is an epistemic hierarchy that provides clear guidelines about why certain data and methods should prevail over others. When integration into the universal tree of life is the major aim of the analysis, methods that deviate from this basic goal can be used only (at best) in a supplementary role. Some network methods may in this context be interpreted simply as tools to detect (and then remove) data conflicts, rather than as a means of representing reticulate events in evolutionary history (Morrison, 2010). Methods that attempt to convert evidence of LGT into support for tree patterns also fall into this camp (e.g., Abby et al., 2012). Lying behind the struggle of prokaryote phylogeny to be integrated into universal phylogeny at the levels of method and data is a tightly focused explanation of evolutionary processes and outcomes. If a range of explanations of evolutionary history were permitted, different methods enabling different representational strategies would be required. It is the inability of tree-associated evolutionary explanations to include more than a thin narrative of prokaryote evolution that ultimately accounts for the limited integration of prokaryote phylogeny within universal phylogeny.

4.3. Explanatory integration: focus and scope

> ‘One of the grand missions of systematics is to reconstruct … the great Tree of Life. As difficult as it may be for modern methodologies to reconstruct this history, and as fraught with reticulation, hybridization events, horizontal gene transfer, and other mechanisms that cloud the picture of organismal history, … at the level of populations and species, *there is only one such history*, even when reticulate … *there is no heterogeneity* … because the history has happened only once’ (Edwards, 2009, p. 2; emphases added).

Arguably more contentious than the data- and method-based problems outlined in Sections 4.1 and 4.2 are the problems of bias and circularity in regard to the tree of life model. A tree with a unique ever-bifurcating topology is presumed to exist, methods are devised to find it, and all data contradicting it are carefully neutralized in order to confirm the prior supposition (McInerney et al., 2011). Sophisticated statistical methods, which look for overall signal in the tree markers rather than the right genes, can only generate limited amounts of support for tree signal (Puigbò et al., 2009; Leigh et al., 2011). Even when tree-oriented methods work, therefore, the majority of evolutionary information is subordinated to the minority. This strategy could be considered justifiable if the model underlying the methods is known from other data or reasoning to be the *only* correct one. But it is not at all clear this is the case (because of all the LGT, endosymbiosis, hybridization and other fusions that have occurred in evolutionary history), and some prokaryote phylogeneticists have argued strongly that the circularity of justifying the choice of tree patterns via the detection of tree patterns is a crucial flaw in efforts to integrate prokaryote phylogeny into universal phylogeny (Doolittle & Bapteste, 2007). Universal phylogeny in its traditional formulation is not about everything that happens in evolutionary history, but a pared-down abstraction of a particular process that explains the bifurcating pattern: vertical descent with modification. Since the pattern of bifurcation is what needs to be explained, and vertical descent with modification (and subsequent divergence) the expected explanation, their decoupling casts light on the restrictions placed around the phenomena to be explained. The failure of prokaryote phylogeny to produce a unique tree of bifurcating branches is not only attributable to its limited explanatory scope, but also the potential misidentification of the evolutionary phenomena to be explained (i.e., tree patterns) in the first place (Doolittle & Bapteste, 2007).

There are several explanatory aspects of universal phylogeny that need revision in light of prokaryote evolution. What explains non-bifurcating patterns? Not just the occurrence of LGT, but the very mode in which prokaryotes speciate (recall here that species are the basic units of universal phylogeny). Speciation is not an abrupt process for any organism, of course, but in prokaryotes it often occurs in such a piecemeal way that it can be conceptualized at best as ‘fragmented’. If only parts of genomes ‘speciate’, and other parts continue independently to recombine, with divergence (splitting) continuing in the 'speciated' part, then the concept of speciation needs wholesale revision for such organisms (Lawrence & Retchless, 2010). This gets to the heart of traditional phylogeny, in which a bifurcation represents a speciation event. The cohesion of many eukaryote genomes, whereby different organisms either can or cannot interbreed as wholes, simply does not apply to partial speciation processes (Retchless & Lawrence, 2010). So in this light, LGT is only a secondarily confounding issue: no such thing as the gene-based tree of life (normally understood as a series of bifurcating lineages) could exist for prokaryotes. For those who think, 'but there *is* a tree of organisms', a non-DNA-based method needs to be found to track those cells, and a model devised that can specify precisely when in evolutionary history those groups of cells formed species.

Some commentators have argued there is a metaphysical fact of the tree of life but that it might not be knowable (Dawkins, in the paraphrased quote in Section 3, assumes both the metaphysical fact and its knowability). However, inability to know the tree is commonly seen as just a temporary epistemic constraint. New methods are expected to reveal more of the unique history of life. But even this 'metaphysical' uniqueness of the tree is questionable. It presumes clear speciation, if it is a tree of species; if it relies on cells always bifurcating, the same fundamental problems apply as for a tree of all species because those cells must be grouped together in bifurcating lineages. Numerous non-bifurcating processes have occurred throughout evolutionary history and contributed to the biodiversity apparent today (for which the tree of life is purportedly explanatory). This plurality of biological facts of the evolutionary matter overwhelms any metaphysical notion of a single tree being able to rule and bind all evolutionary knowledge. At best, universal phylogeny will capture a narrow and abstracted trend in that panoply of processes.

4.4. Integration in universal phylogeny

When summing up the situation of prokaryote phylogeny in relation to universal phylogeny, we should recollect that non-bifurcating processes occur in eukaryotes too, and not just in regard to cellular sub-compartments (mitochondria and plastids). Eukaryote genomes are themselves 'genetic chimeras' (Martin, 2011; Pisani et al., 2007; Koonin, 2010), because they possess informational genes that are very similar to those of archaea, and operational genes that are much closer to those of bacteria. They are the products of genome fusion, but the exact sources of these fusions are contested. In addition, several functional systems in eukaryotes are singly and jointly contributed by bacteria and archaea (e.g., RNA interference machinery), making the evolutionary origins of the eukaryote cells very difficult to trace (Koonin, 2010; O'Malley, 2010b). One way to deal with such chimerism is to worry about the bifurcatory logic of the tree only later on in eukaryote evolution. In other words, follow Ernst Mayr's directives and focus mostly on large sexually reproducing animals, as well as any plants that fit this model (Mayr, 1982; O'Malley, 2010a). Obviously, this strategy would discard a large amount of evolutionary history, including major events in animal evolution, such as the origin of mammals being co-eval with the acquisition of certain viral genes (Cornelis et al., 2012; Mallet et al., 2004). Nevertheless, the segregation of prokaryotic and eukaryotic modes of evolution could have the positive consequence of different representations being developed for non-tree-like and tree-like evolutionary processes.

But what this strategy produces is a paradoxical state of affairs: phylogenetic findings that are not integrated into 'universal' phylogeny because that so-called universality is restrictive. Numerous data, methods and explanations have to be excluded (with choices in one domain having ramifications for the others) in order to make the universal tree hold firm. Justifying what to include and what not to is simply a fact of doing science. But with a focus on integration, the exclusion of certain data, methods and models comes to the fore, and the epistemic and practical value of such selections has to be compared to a more encompassing rationale. All the problems of integration for prokaryote phylogeny seem to exist in eukaryote phylogeny, even animal phylogeny. But in animal phylogeny, the model trumps the data, the methods serve the model, the tree structure remains dominant, incongruence is explained as methodological artefact, and more data argued to be the resolving force of any persistently obscure parts of the tree. It is mostly in prokaryote phylogeny that incongruence is theorized as saying something fundamental about evolutionary process, and about human and technical capabilities for knowing deep evolutionary history.

5. The scope of integrative practices and their normative implications

There are thus two ways in which the failure of integration of prokaryote phylogeny into universal phylogeny can be understood. The first is as a straightforward defeat: certain prokaryote data are recalcitrant to standard phylogenetic methods (i.e., because these data persistently produce incongruence), and explanations of prokaryote evolution require different accounts than those used for the rest of the evolutionary history depicted in universal phylogeny (Bapteste & Burian, 2010; Lapointe et al., 2010). While this situation might be remedied in the future, and many prokaryote phylogeneticists still believe it will be, trends in the field indicate that solutions will be hard to achieve. But this failure of integration is not one that is without any parallels in eukaryote phylogeny, and the same epistemic strategies that attend integration in regard to prokaryote phylogeny vis-à-vis universal phylogeny also attend animal phylogeny. The assumption of a universal tree of life guides and legitimizes the way integration works – not always by greater inclusiveness.

The second diagnosis is more positive: prokaryote evolution has theoretical, methodological, and data-based reasons *not* to be integrated into universal phylogeny. The failure of integration into the universal tree is because of an overriding concern to integrate *more* data (not just the tree-fitting data), *more* methods (not just bifurcatory tree-building methods), and *more* explanatory scope (not just evolutionary processes that can be represented as trees). Why integration fails is because larger scale integration cannot be achieved in the context of the restricted framework of universal phylogeny. We might therefore reasonably conclude that the failure of prokaryote phylogeny to be assimilated within a universal tree is in fact a triumph of integration: in the name of a bigger integrative aim (i.e., that of representing a greater amount of evolutionary history), more limited integration (i.e., in which only bifurcating events are considered, whether prokaryote or eukaryote) has been rejected.

The situation in universal phylogeny can be interpreted, therefore, as one in which a broader view of integration functions as both a positive and negative heuristic. On the negative side, it shows where integration has not and perhaps cannot be achieved. Yet, because of the positive dynamic of integrative practices, guided by a motivation that different activities can be extended and connected, the limits of integration become themselves highly informative. In this sense, integration works as a meta-heuristic along the lines described by Bill Wimsatt in relation to reduction (2006). Whether or not reduction works matters less than what is learned in the process of trying to make it work (and fits the contemporary philosophy of science focus of studying practice rather than outcome). In prokaryote phylogeny conceived as contributing to universal phylogeny, the repeated failure to achieve such integration has turned some microbial evolutionists to the task of detecting and representing horizontal events on the basis of a multiprocess explanation of evolutionary patterns (e.g., Beiko, 2010; Beauregard-Racine et al., 2011).

Elsewhere, I have argued that the tree of life itself is best understood as a heuristic (O'Malley & Koonin, 2011). This diagnosis uses a standard notion of heuristic (i.e., some sort of exploratory conceptual tool), whereas integration (or reduction) is a whole collection of strategies and thus deserving of the label 'meta-heuristic'. As Wimsatt (2006) notes, the systematic failure of such strategic or methodological heuristics leads to the ability to predict the circumstances in which a particular heuristic will fail and how it can be improved. We can see to a large extent how this occurs in phylogeny in relation to the tree of life heuristic, but the situation is less clear in regard to the more abstracted meta-heuristic of integration. The systematic failures of prokaryote phylogeny to fit neatly within universal phylogeny have implications for phylogeny generally, in regard to the limits of tree assumptions and where there is scope for more integrative modelling and methods, and more extensive data synthesis (Bapteste & Burian, 2010). However, the failure of prokaryote phylogeny to integrate in full does not delegitimize tree-of-life-based phylogeny; it simply demarcates it. When those limits are clear, predictions can be made about where to make more inclusive efforts to analyse and represent evolution, so that they encompass non-tree-like data, methods that go beyond tree building, and broader explanations of evolutionary process and pattern. And even more inclusively, thinking about integration as a heuristic strategy has implications for what we consider the norms of science.

Using the notion of integration as a regulative ideal, Grantham (this issue) has suggested that making two-way connections between fields has often been a characteristic of developing fields such as molecular biology. Discounting 'negative integration', or the mere removal of tensions between fields, Grantham focuses on ‘positive integration’, in which densely interconnected webs of practice define and develop fields. My suggestion of integration as a meta-heuristic has some similarities with its characterization as a regulative ideal, but I have focused less on fields and general ‘approaches’ than on specific sets of practice: data, methods, and explanation. Viewing integration as a meta-heuristic means that the focus of philosophical attention is the ongoing process of practice rather than specific aims and outcomes. The various contexts of integration matter a great deal, of course (e.g., between fields, as in Grantham’s (2004a) account; or from context to context, including the context of translation from basic to applied science, as in Leonelli’s account, this issue). However, understanding the contextual relevance of integration is not sufficient for in-depth insight into how integration is achieved, and what prevents it being achieved. Criteria of successful integration are unlikely to be devised because of the very specificities of any integrative activities, but the more commitment there is to multidimensional models of biological processes, the more likely integration is to be valued, aimed for, and acted upon. In this sense, therefore, the notions of meta-heuristic and regulative ideal share similar normative commitments: that integration is required in order to accomplish broader-scope biology, in which multilevel accounts of biological process are increasingly seen as necessary to good science (Wolkenhauer & Hofmeyr, 2007; Brigandt and Love, 2012). When integration of a particular scope is not successful, as in the case of prokaryote phylogeny being integrated into universal phylogeny, we can see highly explanatory reasons for why this is not achieved, how it can lead to broader integrative projects, and yet why a lesser degree of integration might still be considered desirable.

As the ongoing efforts to integrate prokaryote phylogeny into universal phylogeny demonstrate, integration does not always mean greater inclusiveness of data, methods or explanation (indications from the opening quotes to this paper notwithstanding). Integration may involve considerable exclusiveness to achieve the desired integrative aim. The more philosophers discuss integration and how it works in practice, the more likely it is that integration is found to have a variety of aims and criteria of success. But the overarching goal is for some form of integration, and analyses of practice are needed to show how this achievement occurs in different contexts. Examining integrative biology from the pragmatic philosophical point of view recommended by several contributors to this special issue (e.g., Mitchell, this issue; Love, this issue; Bechtel, this issue) requires a multidimensional account of different kinds of practice and how these are synthesized or not synthesized to produce new biological knowledge and research capabilities. As this study of phylogeny has shown, a great deal can be learned from trying to understand integration in action, and not all of it consolidates the assumptions with which my inquiry began – that integration is all about expansion and more in-depth knowledge. Moreover, the interplay between integration and other norms in the life sciences (e.g., reduction, innovation, generality, precision) needs more attention in order to understand better the dynamics of scientific practice. With the burgeoning of integrative biology, philosophers have a great deal of material with which to develop their accounts of integration and how it succeeds and fails.

Acknowledgements

I am very grateful for perspicuous comments and advice from Eric Bapteste, Sara Green, an anonymous referee, and the guest-editor of this special issue on integration, Ingo Brigandt. My research is funded by the Australian Research Council in the form of a Future Fellowship, with additional support from the University of Sydney.

References

Abby, S. S., Tannier, E., Gouy, M., & Daubin, V. (2012). Lateral gene transfer as a support for the tree of life. *Proceedings of the National Academy of Sciences USA*, *109*, 4962-4967.

Aderem, A. (2005). Systems biology: its practice and challenges. *Cell*, *121*, 511-513.

Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud’homme, B., & de Rosa, R. (2000). The new animal phylogeny: reliability and implications. *Proceedings of the National Academy of Sciences USA*, *97*, 4453-4456.

Andam, C. P., Williams, D., & Gogarten, J. P. (2010). Biased gene transfer mimics patterns created through shared ancestry. *Proceedings of the National Academy of Sciences USA*, *107*, 10679-10684.

Anderson, E.S. (1968). The ecology of transferable drug resistance in the enterobacteria. *Annual Review of Microbiology*, *22*, 131-180.

Andersson, J. O. (2009). Horizontal gene transfer between microbial eukaryotes. In Gogarten, M. B., Gogarten, J. P., & Olendzenski, L. C. (Eds.), *Horizontal gene transfer: genomes in flux* (pp. 473-487). NY: Humana.

Archibald, J. M. (2011). Origin of eukaryotic cells: 40 years on. *Symbiosis*, *54*, 69-86.

Archibald, J. M. (2012). Plastid origins. In C. F. Bullerwell (Ed.), *Organelle genetics* (pp. 19-38). Berlin: Springer-Verlag.

Arnold, M. L. (2007). *Evolution through genetic exchange*. Oxford University Press.

Assis, L. C. S. (2009). Coherence, correspondence, and the renaissance of morphology in phylogenetic systematics. *Cladistics*, *25*, 528-544.

Auffray, C., Imbeaud, S., Roux-Rouquié, M., & Hood, L. (2003). From functional genomics to systems biology: concepts and practices. *Comptes Rendus Biologies*, *326*, 879-92.

Bapteste, E., & Burian, R. M. (2010). On the need for integrative phylogenomics, and some steps towards its creation. *Biology and Philosophy*, *25*, 711-736.

Bapteste, E., O'Malley, M. A., Beiko, R. M., Ereshefsky, M., Gogarten, J. P., Franklin-Hall, L., Lapointe, F. J., Dupré, J., Dagan, T., Boucher, Y., & Martin, W. (2009). Prokaryote evolution and the tree of life are two different things. *Biology Direct*, *4*:34, doi:10.1186/1745-6150-4-34.

Bapteste, E., Susko, E., Leigh, J., Ruiz-Trillo, I., Bucknam, J., & Doolittle, W. F. (2008). Alternative methods for concatenation of core genes indicate a lack of resolution in deep nodes of the prokaryotic phylogeny. *Molecular Biology and Evolution*, *25*, 83-91.

Baroni, M, Semple, C., & Steel, M. (2006). Hybrids in real time. *Systematic Biology*, *55*, 46-56.

Bateman, R. M., Hilton, J., & Rudall, P. J. (2006). Morphological and molecular phylogenetic context of the angiosperms: contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *Journal of Experimental Botany*, *57*, 3471-3503.

Beauregard-Racine, J., Bicep, C., Schliep, K., Lopez, P., Lapointe, F.-J., & Bapteste, E. (2011). Of woods and webs: possible alternatives to the tree of life for studying genomic fluidity in *E. coli*. *Biology Direct*, *6*:39, doi: 10.1186/1745-6150-6-39.

Bechtel, W. (1993). Integrating sciences by creating new disciplines: the case of cell biology. *Biology and Philosophy, 8*, 277-99.

Bechtel, W. (this issue). From molecules to clinics: chronobiology as integrative pursuit. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Beiko, R. G. (2010). Gene sharing and genome evolution: networks in trees and trees in networks. *Biology and Philosophy*, *25*, 659-673.

Beiko, R. G. (2011). Telling the whole story in a 10,000-genome world. Biology Direct, 6:34, doi:10.1186/1745-6150-6-34

Beiko, R. G., Harlow, T. J., & Ragan, M. A. (2005). Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences USA, 102*, 14332-14337.

Boucher, Y., & Bapteste, E. (2009). Revisiting the concept of lineage in prokaryotes: a phylogenetic perspective. *BioEssays*, *31*, 526-536.

Bourlat, S. J., Nielsen, C., Economou, A. D., & Telford, M. J. (2008). Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Molecular Phylogenetics and Evolution*, *49*, 23-31.

Boussau, B., Guéguen, L., & Gouy, M. (2008). Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of aquificales in the phylogeny of Bacteria. *BMC Evolutionary Biology*, *8*:22, doi: 10.1186/1471-2148-8-272.

Brigandt, I. (2010). Beyond reductionism and pluralism: toward an epistemology of explanatory integration. *Erkenntis, 73*, 295-311.

Brigandt, I. (this issue). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Brigandt, I. (forthcoming). Intelligent design and the nature of science: philosophical and pedagogical points. In Kampourakis, K. (Ed.), *Philosophical Issues In Biology Education*. NY: Springer.

Brigandt, I., & Love, A. C. (2012). Reductionism in biology. In Zalta, E. N. (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/sum2012/entries/reduction-biology/

Brochier, C., Philippe, H., Moreira, D. (2000). The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. *Trends in Genetics*, *16*, 529-533.

Budd, A. F., & Stolarski, J. (2010). Corallite wall and septal microstructure in scleractinian reef corals: comparison of molecular clades within the family Faviidae. *Journal of Morphology*, *272*, 66-88.

Burian, R. M. (1993). Unification and coherence as methodological objectives in the biological sciences. *Biology and Philosophy, 8*, 301-318.

Bybee, S. M., Zaspel, J. M., Beucke, K. A., Scott, C. H., Smith, B. W., & Branham, M. A. (2010). Are molecular data supplanting morphological data in modern phylogenetic studies? *Systematic Entomology*, *35*, 2-5.

Cali, A., Calvanese, D., de Giacomo, G., & Lenzerini, M. (2003). Data integration under integrity constraints. *Information Systems, 29*, 147-63.

Caterino, M. S., Cho, S., & Sperling, F. A. H. (2000). The current state of insect molecular systematics: a thriving tower of Babel. *Annual Review of Entomology*, *45*, 1-54.

Chuang, H.-Y., Hofree, M., & Ideker, T. (2010). A decade of systems biology. *Annual Review of Cell and Developmental Biology*, *26*, 721-744.

Cohen, O., Gophna, U., & Pupko, T. (2011). The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Molecular Biology and Evolution*, *28*, 1481-1489.

Cornelis, G., Heidmann, O., Bernard-Stoecklin, S., Reynaud, K., Véron, G., Mulot, B., Dupressoir, A., & Heidmann, T. (2012). Ancestral capture of *syncytin-Car1*, a fusogenic endogenous retroviral *envelope* gene involved in placentation and conserved in Carnivora. *Proceedings of the National Academy of Sciences USA*, *109*, E432-E441.

Coughter, J. P., & Stewart, G. J. (1989). Genetic exchange in the environment. *Antonie van Leeuwenhoek*, *55*, 15-22.

Dagan, T., Artzy-Randrup, Y., & Martin, W. F. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proceedings of the National Academy of Sciences USA*, *105*, 10039-10044.

Dagan, T., & Martin, W. F. (2006). The tree of one percent. *Genome Biology*, *7*:118, doi:10.1186/gb-2006-7-10-118.

Darwin, C. (1872). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (6th ed). London: John Murray.

Dawkins, R. (2003). *The devil's chaplain: reflections on hope, lies, science, and love*. Boston: Houghton Mifflin.

Dayhoff, M. O., Barker, W. C., & McLaughlin, P. J. (1974). Inferences from protein and nucleic acid sequences: early molecular evolution, divergence of kingdoms and rates of change. *Origins of Life*, *5*, 311-330.

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wiezer, A., Bäumer, S., Jacobi, C., et al. [12 others]. (2002). The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *Journal of Molecular Microbiology and Biotechnology*, *4*, 453-61.

DeSalle, R., & Schierwater, B. (2008). An even 'newer' animal phylogeny. *BioEssays*, *30*, 1043-1047.

Didelot, X., and Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. *Trends in Microbiology*, *18*, 315-322.

Donoghue, P. C. J., & Benton, M. J. (2007). Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology and Evolution*, *22*, 424-431.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science*, *284*, 2124-2128.

Doolittle, W. F., & Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proceedings of the National Academy of Sciences USA*, *104*, 2043-2049.

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., et al. [8 others] (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, *452*, 745-749.

Ebersberger, I., de Matos Simoes, R., Kupczok, A., Gube, M., Koteh, E., Voigt, K., & von Haeseler, A. (2011). A consistent phylogenetic backbone for the fungi. *Molecular Biology and Evolution*, doi:10.1093/molbev/msr285.

Eernisse, D. J., & Kluge, A. G. (1993). Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Molecular Biology and Evolution*, *10*, 1170-1195.

Edgecombe, G. D. (2010). Arthropod phylogeny: an overview from the perspectives of morphology, molecular data and the fossil record. *Arthropod Structure and Development*, *39*, 74-87.

Edgecombe, G. D., Giribet, G., Dunn, C. W., Hejnol, A., Kristensen, R. M., Neves, R. C., Rouse, G. W., Worsaae, K., & Sørensen, M. V. (2011). Higher-level metazoan relationships: recent progress and remaining questions. *Organisms, Diversity & Evolution*, *11*, 151-172.

Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution*, *63*, 1-19.

Feil, E.J., Holmes, E.C., Bessen, D.E., Chan, M.S., Day, N.P., Enright, M.C., Goldstein, R., Hood, D.W., Kalia, A., Moore, C.E., Zhou. J., & Spratt, B.G. (2001). Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proceedings of the National Academy of Sciences USA, 98*, 182-187.

Galtier, N., Daubin, V. (2008). Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society London B*, *363*, 4023-4029.

Gee, H. (2003). Ending incongruence. *Nature*, *425*, 782.

Gerson, E. M. (this issue). Integration of specialties: An institutional and organizational view. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Giribet, G. (2010). A new dimension in combining data? The use of morphology and phylogenomic data in metazoan systematics. *Acta Zoologica*, 91, 11-19.

Ghosh, S., Matsuoka, Y., Asai, Y., Hsin, K.-Y., and Kitano, H. (2011). Software for systems biology: from tools to integrated platforms. *Nature Reviews Genetics*, *12*, 821-832.

Gogarten, J. P., Doolittle, W. F., & Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution*, *19*, 2226-2238.

Gophna, U., & Ofran, Y. (2011). Lateral acquisition of genes is affected by the friendliness of their products. *Proceedings of the National Academy of Sciences USA, 108*, 343-348.

Grantham, T. A. (2004a). Conceptualizing the (dis)unity of science. *Philosophy of Science, 71*, 133-55.

Grantham, T. A. (2004b). The role of fossils in phylogeny reconstruction: why is it so difficult to integrate paleobiological and neontological evolutionary biology? *Biology and Philosophy*, *19*, 687-720.

Grantham, T. A. (this issue). Integrative pluralism as a regulative ideal. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Griesemer, J. R. (this issue). Integration practices in a contemporary research system: The case of David Wake's model-taxon based research in evolutionary morphology. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Haggerty, L. S., Martin, F. J., Fitzpatrick, D. A., & McInerney, J. O. (2009). Gene and genome trees conflict at many levels. *Philosophical Transactions of the Royal Society London B*, *364*, 2209-2219.

Halanych, K. M. (2004). The new view of animal phylogeny. *Annual Review of Ecology, Evolution, and Systematics*, *35*, 229-256.

Hermsen, E. J., & Hendricks, J. R. (2008). W(h)ither fossils? Studying morphological character evolution in the age of molecular sequences. *Annals of the Missouri Botanical Garden*, *95*, 72-100.

Hilario, E., & Gogarten, J.P. (1993). Horizontal transfer of ATPase genes – the tree of life becomes a net of life. *BioSystems*, *31*, 111-119.

Hotopp, J. C., Clark, M. E., Oliveira, D. C., Foster, J. M., Fischer, P., Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N., Wang, S., et al. [10 others]. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science*, *317*, 1753-1756.

Huang, D., Licuanan, W. Y., Baird, A. H., & Fukami, H. (2011). Cleaning up the ‘Bigmessidae’: molecular phylogeny of scleractinian corals from Faviidae, Merulinidae, Pectiniidae and Trachyphylliidae. *BMC Evolutionary Biology*, *11*:37, doi:10.1186/1471-2148-11-37.

Huang, J., & Gogarten, J. P. (2006). Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends in Genetics*, *22*, 361-366.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*, 254-267.

Huson, D. H., & Scornavacca, C. (2011). A survey of combinatorial methods for phylogenetic networks. *Genome Biology and Evolution*, *3*, 23-35.

Hyman, A. A. (2011). Whither systems biology. *Philosophical Transactions of the Royal Society London B*, 366, 3635-3637.

Ideker, T., Bafna, V., & Lemberger, T. (2007) Integrating scientific cultures. *Molecular Systems Biology* 3:105, doi: 10.1038/msb4100145.

Jeffroy, O., Brinkmann, H., Delsuc, F., & Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends in Genetics*, *22*, 225-231.

Jenner, R. A. (2004a). Accepting partnership by submission? Morphological phylogenetics in a molecular millennium. *Systematic Biology*, *53*, 333-342.

Jenner, R. A. (2004b). When molecules and morphology clash: reconciling conflicting phylogenies of the Metazoa by considering secondary character loss. *Evolution and Development*, *6*, 372-378.

Jenner, R. A. (2011). Use of morphology in criticizing molecular trees. *Journal of Crustacean Biology*, *31*, 373-377.

Jones, D., & Sneath, P. H. A. (1970). Genetic transfer and bacterial taxonomy. *Bacteriological Reviews*, *34*, 40-81.

Keeling, P. J. (2009). Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Current Opinion in Genetics and Development*, *19*, 613-619.

Keeling P. J., & Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics, 9*, 605-618.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48*: 507-31.

Kitcher, P. (1999). Unification as a regulative ideal. *Perspectives on Science, 7*, 337-348.

Klenk, H.-P., & Göker, M. (2010). En route to a genome-based classification of Archaea and Bacteria? *Systematic and Applied Microbiology*, *33*, 175-182.

Koonin, E. V. (2010). The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biology*, 11:209, doi:10.1186/gb-2010-11-5-209.

Lapierre, P., & Gogarten, J. P. (2008). Estimating the size of the bacteria pan-genome. *Trends in Genetics*, *25*, 107-110.

Lapointe, F.-J., Lopez, P., Boucher, Y., Koenig, J., & Bapteste, E. (2010). Clanistics: a multi-level perspective for harvesting unrooted gene trees. *Trends in Microbiology*, *18*, 341-347.

Lauffenburger, D. A. (2012). The multiple dimensions of *Integrative Biology*. *Integrative Biology*, *4*, 9.

Lawrence, J. G. (2002). Gene transfer in bacteria: speciation without species. *Theoretical Population Biology*, *61*, 449-460.

Lawrence, J. G., & Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences USA*, *95*, 9413-9417.

Lawrence, J. G., & Retchless, A. C. (2009). The myth of bacterial species and speciation. *Biology and Philosophy*, *25*, 569-588.

Lecointre, G., & Deleporte, P. (2005). Total evidence requires exclusion of phylogenetically misleading data. *Zoologica Scripta*, *34*, 101-117.

Leigh, J. W., Lapointe, F.-J., Lopez, P., and Bapteste, E. (2011). Evaluating phylogenomic congruence in the post-genomic era. *Genome Biology and Evolution*, *3*, 571-587.

Lenzerini, M. (2002). Data integration: a theoretical perspective. *PODS '02, Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (pp. 233-246). NY: ACA.

Leonelli, S. (2008). Bio-ontologies as tools for integration in biology. *Biological Theory, 3*, 7-11.

Leonelli, S. (this issue). What counts as data in integrative biology?. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Leplae, R., Hebrant, A., Wodak, S. J., and Toussaint, A. (2004). ACLAME: A CLAssification of Mobile genetic Elements. *Nucleic Acids Research*, *32 (Database issue)*, D45-D49.

Linder, C. R., & Rieseberg, L. H. (2004). Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany*, *91*, 1700-1708.

Liu, E. T. (2005). Systems biology, integrative biology, predictive biology. *Cell*, *121*, 505-506.

Loftus, B., Anderson, I., Davies, R., Alsmark, U.C., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R.P., Mann, B.J., et al. [44 others] (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature*, *433*, 865-868.

Lockhart, P. J., & Cameron, S. A. (2001). Trees for bees. *Trends in Ecology and Evolution*, *16*, 84-88.

Love, A. C. (this issue). Dimensions of integration in historical explanation: physics, genetics, and the origins of novelty. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Lukjancenko, O., Wassenaar, T. M., & Ussery, D. W. (2010). Comparison of 61 sequenced Escherichia coli genomes. *Microbial Ecology*, *60*, 708-720.

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L, & Mandrand, B. (2004). The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. *Proceedings of the National Academy of Sciences USA*, *101*, 1731-1736.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, *20*, 229-237.

Martin, W. F. (1996). Is something wrong with the tree of life? *BioEssays*, *18*, 523-527.

Martin, W. F. (1999). Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. B*ioEssays*, *21*, 99-104.

Martin, W. F. (2011). Early evolution without a tree of life. *Biology Direct*, *6*:36, doi:10.1186/1745-6150-6-36.

Martin, W. F., & Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiology*, *118*, 9-17.

Maynard Smith, J., Feil, E. J., & Smith, N. H. (2000). Population structure and evolutionary dynamics of pathogenic bacteria. *BioEssays*, *22*, 1115-1122.

Mayr, E. (1982). *The growth of biological thought: diversity, evolution, and inheritance*. Cambridge, MA: Harvard University Press.

McInerney, J. O., Cotton, J. A., & Pisani, D. (2008). The prokaryotic tree of life: past, present … and future? *Trends in Ecology and Evolution*, *23*, 276-281.

McInerney, J. O., Pisani, D., Bapteste, E., & O'Connell, M. J. (2011). The public goods hypothesis for the evolution of life on Earth. *Biology Direct*, *6*:41, doi:10.1186/1745-6150-6-41.

Medini, D., Donati, C., Tettelin, H., Masignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics and Development*, *1*, 589-594.

Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Mitchell, S. D. (this issue). Integrative strategies in explanations of protein folding. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Morrison, D. A. (2010). Using data-display networks for exploratory data analysis in phylogenetic studies. *Molecular Biology and Evolution*, *27*, 1044-1057.

Morrison, M. (2000). *Unifying scientific theories: physical concepts and mathematical structures*. Cambridge: Cambridge University Press.

Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., & Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science*, *324*, 1724-1726.

Mykles, D. L., Ghalambor, C. K., Stillman, J. H., & Tomanek, L. (2010). Grand challenges in comparative physiology: integration across disciplines and across levels of biological organization. *Integrative and Comparative Biology*, *50*, 6-16.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R,J,, Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K,A,, et al. [19 others] (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, *399*, 323-329.

O'Malley, M. A. (2010a). Ernst Mayr, the tree of life, and philosophy of biology. *Biology and Philosophy*, *25*, 529–552.

O'Malley, M. A. (2010b). The first eukaryote cell: an unfinished history of contestation. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, *41*, 212–224.

O'Malley, M. A., & Boucher, Y. (2005). Paradigm change in evolutionary microbiology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 183-208.

O'Malley, M. A., and Koonin, E. V. (2011). How stands the Tree of Life a century and a half after *The Origin*? *Biology Direct*, 6:32, doi:10.1186/1745-6150-6-32.

O'Malley, M. A., & Soyer, O. S. (2012). The roles of integration in molecular systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 58-68.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In Feigl, H., Scriven, M., & Maxwell, G. (Eds.), *Minnesota Studies in Philosophy of Science,* 2, 3-36, Minneapolis: Minnesota University Press.

Paps, J., Baguñà, J., & Riutort, M. (2009). Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. *Proceedings of the Royal Society of London B*, *276*, 1245-1254.

Patel, M., & Nagi, S. (2010). *The role of model integration in complex systems modelling: an example from cancer biology*. Berlin: Springer.

Pauling, L., & Zuckerkandl, E. (1963). Chemical paleogenetics: molecular 'restoration studies' of extinct forms of life. *Acta Chemica Scandinavica*, *17*, S9-S16.

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wöheide, G., & Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, *9* (3), e1000602.

Pisani, D., Cotton, J. A., & McInerney, J. O. (2007). Supertrees disentangle the chimerical origins of eukaryotic genomes. *Molecular Biology and Evolution*, *24*, 1752-1760.

Powell, A., O'Malley, M. A., Müller-Wille, S. E. W., Calvert, J., and Dupré, J. (2007). Disciplinary baptisms: a comparison of the naming stories of genetics, molecular biology, genomics and systems biology. *History and Philosophy of the Life Sciences*, *29*, 5–32.

Prill, R. J., Marbach, D., Saez-Rodriguez, J., Sorger, P. K., Alexopoulos, L. G., Xue, X., Clarke, N. D., Altan-Bonnet, G., & Stolovitzky, G. (2010). Towards a rigorous assessment of systems biology models: the DREAM3 challenge. *PLoS One*, *5*(2), e9202.

Puigbò, P., Wolf, Y. I., & Koonin, E. V. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *Journal of Biology*, *8*:59, doi:10.1186/jbiol159.

Ragan, M. A., & Beiko, R. G. (2009). Lateral genetic transfer: open issues. *Philosophical Transactions of the Royal Society London B*, *364*, 2241-2251.

Ragan, M. A., Harlow, T. J., & Beiko, R. G. (2006). Do different surrogate methods detect lateral genetic transfer events of different relative ages? *Trends in Microbiology*, *14*, 4-8.

Reanney, D. (1977). Gene transfer as a mechanism of microbial evolution. *BioScience*, *27*, 340-344.

Regier, J. C., Shultz, J. W., Ganley, A R. D., Hussey, A., Shi, D., Ball, B., Swick, A., Stajich, J. E., Cummings, M. P., Martin, J. W., & Cunningham, C. W. (2008). Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Systematic Biology*, *57*, 920-938.

Retchless, A. C., & Lawrence, J. G. (2010). Phylogenetic incongruence arising from fragmented speciation in enteric bacteria. *Proceedings of the National Academy of Sciences USA*, *107*, 11453-11458.

Richards, T. A., Soanes, D. M., Foster, P. G., Leonard, G., Thornton, C. R., & Talbot, N. J. (2009). Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell*, *21*, 1897-1911.

Ripoll, C, Guespin-Michel, J., Norris, V., & Thellier, M. (1998). Defining integrative biology. *Complexity*, *4*, 19-20.

Rokas, A., & Carroll, S. B. (2006). Bushes in the tree of life. *PLoS Biology*, *4* (11), e352.

Roth, K. J. (1994). Second thoughts about interdisciplinary studies. *American Educator*, *Spring*, 44-48.

Schaffner, K. F. (1993). Theory structure, reduction, and disciplinary integration in biology. *Biology and Philosophy*, *8*, 319-347.

Schilling, M., Pfeifer, A. C., Bohl, S., & Klingmüller, U. (2008). Standardizing experimental protocols. *Current Opinion in Biotechnology*, *19*, 354-359.

Schwartz, R. M., & Dayhoff, M. O. (1978). Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts: A perspective is derived from protein and nucleic acid sequence data. *Science*, *199*, 395-403.

Scotland, R. W., Olmstead, R. G., & Bennett, J. R. (2003). Phylogeny reconstruction: the role of morphology. *Systematic Biology*, *52*, 539-548.

Sidow, A., & Wilson, A.C. (1990). Compositional statistics: An improvement of evolutionary parsiomony and its application to deep branches in the tree of life. *Journal of Molecular Evolution, 31*, 51-68.

Stanier, R. Y., & van Niel, C. B. (1962). The concept of a bacterium. *Archiv für Mikrobiologie, 42*, 17-35.

Suárez-Diáz, E., & Anaya-Muñoz, V. H. (2008). History, objectivity, and the construction of molecular phylogenies. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *39*, 451-468.

Sullivan, D. E., Gabbard, J. L. Jr., Shukla, M., & Sobrai, B. (2010). Data integration for dynamic and sustainable systems biology resources: challenges and lessons learned. *Chemical Biodiversity*, *7*, 1124-1141.

Telford, M. J. (2008). Resolving animal phylogeny: a sledgehammer for a tough nut? *Developmental Cell*, *14*, 457-459.

Telford, M. J., & Copley, R. R. (2011). Improving animal phylogenies with genomic data. *Trends in Genetics*, *27*, 186-195.

Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: The bacterial pan-genome**.** *Current Opinion in Microbiology*, *12*, 472-477.

Vriesendorp, B., & Bakker, F. T. (2005). Reconstructing patterns of reticulate evolution in angiosperms: what can we do? *Taxon*, *54*, 593-604.

Wake, M. H. (2003). What is ‘integrative biology’? *Integrative and Comparative Biology*, *43*, 239-241.

Wake, M. H. (2008). Integrative biology: science for the 21st century. *BioScience*, *58*, 349-53.

Welch, J. J., & Bromham, L. (2006). Molecular dating when rates vary. *Trends in Ecology and Evolution*, *20*, 320-327.

Welch, R. A., Burland, V., Plunkett, G. 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A., Hackett, J., et al. [9 others]. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, *99*, 17020-17024.

Wheelis, M. L., Kandler, O., & Woese, C. R. (1992). On the nature of global classification. *Proceedings of the National Academy of Sciences USA*, *89*, 2930-2934.

Whittaker, R. H. (1969). New concepts of kingdoms of organisms. *Science, 163*, 150-160.

Wiens, J. J., Kuczynski, C. A., Townsend, T., Reeder, T. W., Mulcahy, D. G., & Sites, J. W. Jr. (2010). Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. *Systematic Biology*, *59*, 674-688.

Wilson, A. C., Carlson, S. S., & White, T.J. (1977). Biochemical evolution. *Annual Review of Biochemistry*, *46*, 573-639.

Wimsatt, W. C. (2006). Reductionism and its heuristics: making methodological reductionism honest. *Synthese*, *151*, 445-475.

Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews*, *51*, 221-271.

Woese, C. R. (1996). Phylogenetic trees: whither microbiology? *Current Biology, 6*, 1060-1063.

Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences USA*, *97*, 8392-8396.

Woese, C.R. (2005). The archaeal concept and the world it lives in: a retrospective. In Govindjee, Beatty, J. T., Gest, H., & Allen, J.F. (Eds.), *Discoveries in photosynthesis* (pp. 1109-1120). Dordrecht: Springer.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences USA*, *74*, 5088-5090.

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences USA*, *87*, 4576-4579.

Wolkenhauer, O., & Hofmeyr, J.-H. S. (2007). An abstract cell model that describes the self-organization of cell function in living systems. *Journal of Theoretical Biology*, *246*, 461-476.

Woolley, S. M., Posada, D., & Crandall, K. A. (2008). A comparison of phylogenetic network methods using simulation. *PLoS One*, *3*(4), e1913, doi:10.1371/journal.pone.0001913.

Wortley, A. H., & Scotland, R. W. (2006). The effect of combining molecular and morphological data in published phylogenetic analyses. *Systematic Biology*, *55*, 677-685.

Wylie, A. (1999). Rethinking unity as a ‘working hypothesis’ for philosophy of science: how archaeologists exploit the disunities of science. *Perspectives on Science, 7*, 293-317.

Xie, J., Fu, Y., Jiang, D., Li G., Huang, J., Li, B., Hsiang, T., & Peng, Y. (2008). Intergeneric transfer of ribosomal genes between two fungi. *BMC Evolutionary Biology*, *8*:87, doi:10.1186/1471-2148-8-87.

Yap, W. H., Zhang, Z., & Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *Journal of Bacteriology*, *181*, 5201-5209.

Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., & Noll, K.M. (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proceedings of the National Academy of Sciences USA, 106*, 5865-5870.