Brit. J. Phil. Sci. **59** (2008), 745–765

Evolution without Species: The Case of Mosaic Bacteriophages

Gregory J. Morgan and W. Brad Pitts

ABSTRACT

Recent work in viral genomics has shown that bacteriophages exhibit a high degree of mosaicism, which is most likely due to a long history of prolific horizontal gene transfer (HGT). Given these findings, we argue that each of the most plausible attempts to properly classify bacteriophages into distinct species fail. Mayr's biological species concept fails because there is no useful viral analog to sexual reproduction. Phenetic species concepts fail because they obscure the mosaicism and the rich reticulated viral histories. Phylogenetic species concepts, even when extended to take into account reticulation, fail because there is no non-arbitrary distinction between recombination events that create a new viral species and those that do not. There is good reason to think that bacteriophages, arguably the Earth's most abundant biological agent, evolve without forming species.

1 Introduction

It is commonly thought that the existence of evolution by natural selection entails the formation and persistence of species. Charles Darwin, in his *On the Origin of Species* argues that it 'follows inevitably' from the struggle for existence

doi:10.1093/bjps/axn038

that 'varieties' become 'good and distinct species' (Darwin [1859], p. 61). Others, while not putting the relationship between natural selection and species so starkly, endorse the view that species are *the* units of evolution (Eldredge and Cracraft [1980]; Wiley [1981]). Even most opponents of the view that species are the unit of evolution concede that species are *a* unit of selection—in addition they argue that higher taxa can also be units of evolution. (See Ereshefsky [1991] for a review). Contrary to this standard view, we will argue that prolific horizontal gene transfer (HGT) among bacteriophages and the resulting pervasive mosaicism render all current species concepts poorly suited to these viruses, and that, as a result, one cannot properly classify what is arguably the world's most abundant evolving biological agent into distinct species. We will further suggest, albeit briefly, that the role viruses play in facilitating HGT among their hosts causes similar, yet less severe, mosaicism in viral hosts.

The argument presented here runs counter to the more traditional approach of the International Committee on the Taxonomy of Viruses (ICTV) that attempts to assign every virus a species, a genus, a family, and an order (Fauquet et al. [2005]). For example, the well-studied bacteriophage λ is said to be a member of the species '*Enterobacteria phage* λ,' the genus '*λ-like phages*,' the family '*Siphoviridae*' (phages with long non-contractile tails), and the order '*Caudivirales*' (tailed phages). To be fair, in the latest edition of *Virus Taxonomy*, the ICTV do acknowledge that due to problems associated with HGT the classification of *Caudivirales* should be considered 'provisional' (Fauquet et al. [2005], p. 41). For reasons that will become apparent below, we think such a hierarchical system (in this case based on tail morphology) obscures the complexity of viral history and diversity and should be abandoned.

We begin with a brief review of the biology behind bacteriophage lifecycles and an explanation of how HGT occurs between different bacteriophage genomes within a co-infected host cell. Then we survey the species concepts that seem to be most applicable to virus taxonomy: phenetic species concepts; Ernst Mayr's biological species concept with an extension for asexual species; phylogenetic species concepts and an extension to deal with reticulated phylogenies; ecological species concepts; and Richard Boyd's homeostatic property cluster theory of natural kinds applied to biological species. In each case, we evaluate the concept as a candidate for use in viral taxonomy and find it wanting. None of these concepts can adequately capture both the pervasive mosaicism and the highly reticulated evolutionary history that results from prolific HGT, or so we argue. We suggest that bacteriophage evolution occurs without well-defined species.

2 The Biology of Viruses

2.1 Bacteriophage life cycles

The viruses that attack bacteria are called bacteriophages (or phages). They are perhaps the most abundant biological agent on Earth. It is estimated that there are 10^{31} individual bacteriophages in the biosphere (Hendrix [2005], p. 1). The gross structure and major details of the life cycles of the double-stranded DNA bacteriophages are relatively well understood (Calendar [2006]). Between host cell infections, an exemplary bacteriophage consists of a polyhedral protein capsid that contains DNA and a cylindrical 'tail' with extending tail fibers and tail spike. In this state, bacteriophages have no motility and no intrinsic ability to perform metabolism or replication. In a solution of bacteria, they drift randomly until their tail fiber proteins happen to encounter a specific bacterial membrane receptor. The attachment phase begins with the adherence of tail fibers to the bacterial cell surface receptors, thereby bonding the phage to the host cell wall. The tail spikes often have enzymes that work to degrade the bacterial cell wall and allow the entry of viral genetic material. Only the viral genetic material enters the cell; the protein capsid, as well as the tail and its fibers, remain attached to the cell wall. At this point, various factors determine whether the virus enters one of two life cycles, *lytic* or *lysogenic*.

In the *lytic* cycle of phage replication, the virus 'hijacks' the metabolism of the host cell and directs the synthesis of new viral particles. Phages that replicate this way are called 'virulent.' Among the first products of the transcribed and translated viral DNA are nucleases, which degrade portions of the host DNA so that the viral genetic material is preferentially replicated, transcribed, and translated. The relatively small size of the viral particles—during the lytic cycle, hundreds of intact phages can be synthesized within a single bacterial cell—requires that their genomes be streamlined. Decoding of the virus DNA results primarily in the construction of the protein capsid (which will contain the replicated viral DNA), the tail unit, and lysozyme to allow the newly replicated viruses to escape the host cell. The lytic cycle ends with a destroyed host bacterium and the release of multiple copies of the replicated virus particle.

In contrast to exclusively lytic phages, many bacteriophages can be *lysogenic*. Phages that enter the lysogenic life cycle are termed 'temperate'. For example, phage λ can enter either life cycle depending on various environmental conditions, such as the physiological state of the host bacterium. After phage λ DNA enters the host cell, a phage-encoded enzyme catalyzes a recombination event between phage and bacterial DNA. The end result of entering the lysogenic cycle is phage DNA integrated into the genome of the host bacterium with

the phage genes, whose expression would lead to lytic growth, kept silent by a repressor protein encoded by a phage gene. This integrated viral DNA is referred to as a 'prophage'. The prophage can be replicated indefinitely as part of the bacterial chromosome and passed vertically to future descendents of the host bacterium. There are conditions, however, that cause the prophage to discontinue its production of the repressor protein. If these conditions obtain, the phage DNA will be excised and will enter the lytic cycle. (For a comprehensive overview of bacteriophage biology, see Calendar [2006]).

2.2 Mechanisms of HGT

While the method by which viral genetic material is vertically propagated from an infecting virus to a newly replicated virus is reasonably well understood, HGT, sometimes called lateral gene transfer, is the principal source of the specific difficulties involved in viral classification. HGT is the exchange of genetic material between different viruses. Although viral HGT mechanisms involve both virulent and temperate phages, HGT occurs only between viruses within a host cell.

There are essentially two means of laterally transferring viral genetic material. First, consider a virulent virus infecting a bacterial cell that already contains a viral genome inserted into the bacterial genome (i.e., a 'prophage'). Among the first actions of the new infecting phage is the production of nucleases to degrade parts of the bacterial chromosome. As the new phage begins construction and maturation of new viruses, there is a possibility that nucleases will release portions of the prophage that are then successfully integrated (i.e., integrated without rendering the phage nonfunctional when it infects another host) into the replicated genome of the infecting phage. If such an infection occurs, it is possible that a novel 'functional' phage that contains new genes from a different phage will be created. Second, consider a similar situation in which a temperate phage is infecting a bacterium that already contains a prophage. It is possible that the infecting phage's DNA integrates successfully *within the existing prophage*. If this pre-existing prophage is then induced to excise, it is possible that a novel phage that contains bacteriophage genes from each of the two phages will be created.

While production of viable phages by either mechanism might seem remote, the viable mosaic viruses studied today are only those that have undergone these processes with successful results; the others will have been eliminated by selective factors before biologists were able to study them.

Given enough time, HGT causes mosaicism in most, if not all, bacteriophages (and to a lesser extent their bacterial hosts). Mosaicism here refers to the chimerical nature of both genomes *and* phenotypic expression among bacteriophages. In bacteriophages, HGT plays the role that point mutation usually

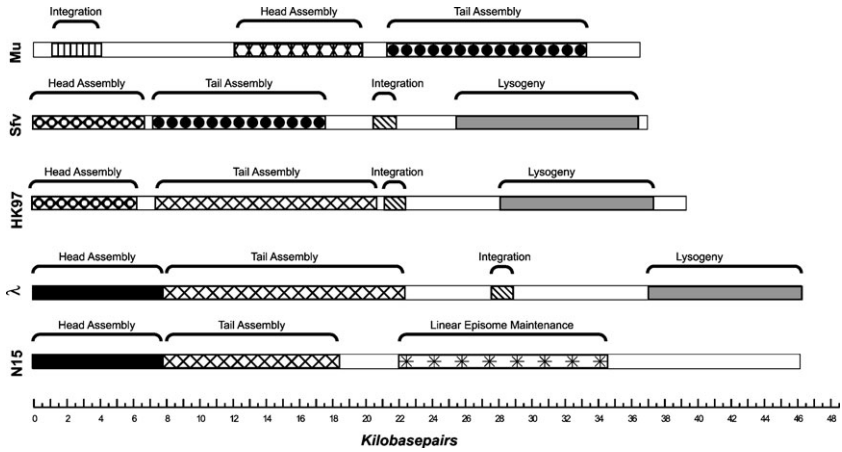

Figure 1. A comparison of five bacteriophage genomes illustrating genetic mosaicism at the level of the genetic module. Each genome contains a number of genetic modules that consist of clustered genes with related functions. Modules that have a number of detectable homologous genes are treated as homologous modules and are depicted with the same pattern. For example, the head assembly module appears homologous in phages N15 and λ, but not in phages N15 and HK97. This diagram does not represent the mosaicism that is also found *within* modules and genes.

plays in the conventional explanation of the generation of biological diversity. The mosaicism is most easily seen in comparisons of two or more bacteriophage genomes: stretches of a genome are very similar and bordering stretches are very different. The mosaicism occurs at the level of the genetic module (Figure 1), at the level of the gene (Figure 2), and within genes. For example, there are phages whose similar capsids suggest they belong to one classificatory group when, in fact, the majority of the genomic comparison indicates two very different evolutionary histories that place them in two different groups. In principle, mosaicism can occur with any number of phage phenotype characteristics (protein capsid, base plates, protein tail fibers, regulatory proteins, etc.). (For more detailed overviews of HGT and bacteriophage evolution, see Hendrix et al. [2000] and Brüssow & Desiere [2006]).

3 The Species Problem and Species Concepts

The *species problem* refers to the question of the ontological status and the correct characterization of biological species. Ernst Mayr ([1996]) emphasizes this duality by distinguishing between the 'species taxon' problem and the 'species category' problem. A species taxon, according to Mayr, is a 'concrete zoological or botanical object consisting of a classifiable population (or group of populations) of organisms. . .' (Mayr [1996], p. 267). *Canis familiaris* is a

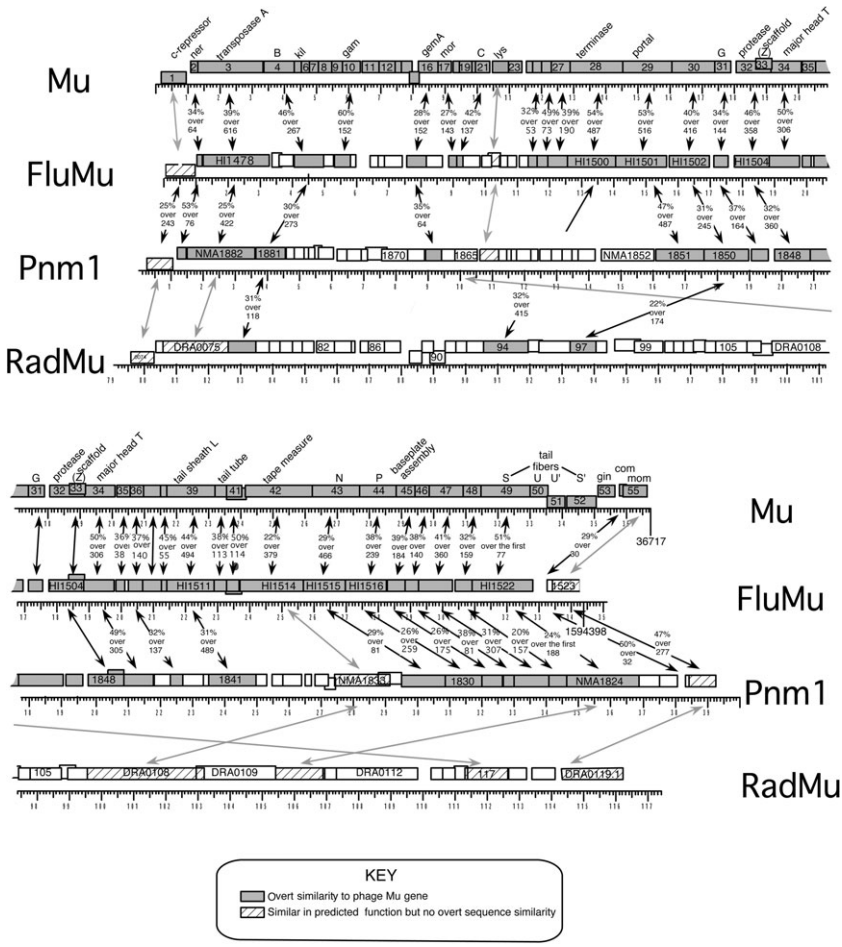

Figure 2. An alignment of four bacteriophage genomes illustrating genetic mosaicism at the level of individual genes. The top genome is the well-studied bacteriophage Mu and the bottom three genomes are prophages found in the genomes of *H. influenzae*, *N. meningiditis* and *D. radiodurans*. The ruler scales are in kilobases of DNA. The rectangular boxes represent putative genes. The black arrows illustrate probable homologies with the degree of similarity given by an alignment of predicted amino acid sequences over a specified number of residues. Gray double-headed arrows point to genes thought to have similar function to a bacteriophage Mu gene but with no significant similarity in the amino acid sequences. Adapted from Morgan et al. [2002].

species taxon. The species category is more abstract; it is 'the rank in Linnaean hierarchy. . .and is defined by the species definition' (Mayr [1996], p. 267). The species definition determines the class of populations that biologists correctly call species. The species category problem is to define what all species taxa have in common. For example, are species taxa classes or individuals? (For the

classic papers on species as individuals, see Ghiselin [1974] and Hull [1978]). The species taxon problem (or problems) is concerned with the identification of the members of a species taxon.

There have been at least several dozen attempts to define the species category. None of the attempts are without problems. Consequently, some philosophers of biology adopt pluralism and argue that we need different concepts in different situations. (See for example Dupré [1993]; Kitcher [1984]; Ereshefsky [2001]). Other philosophers of biology are monists and argue that there is only one correct definition of species. (See for example, Sober [1984]; Ghiselin [1987], [2000]; Hull [1987]). If we are correct about bacteriophage taxonomy, then both groups have a small victory. If no species concept is applicable to bacteriophages, then there is no universal species concept covering all evolving biological agents. On the other hand, it does not follow that we must adopt pluralism because some parts of the biological world are not properly classifiable into species. The remaining classifiable parts might still fall under one species concept.

One way to show that bacteriophages cannot be classified using the various definitions of the species concept is to examine each type of definition. We follow Sterelny and Griffiths ([1999]) in grouping the plethora of species concepts into groups to make discussion of them manageable. Our strategy will be to show how the most plausible groups of species concepts each fail to properly classify bacteriophages. Some of the species concepts use synchronic properties to define a species. Others appeal to diachronic biological history. In either case, prolific HGT or the resulting pervasive mosaicism render these species concepts ill-suited to properly classify bacteriophages, or so we will argue.

3.1 Phenetic species concepts

A phenetic species concept is one in which membership is determined by the relative similarity of the individuals within the group. The similarity can be physical, genetic, or even behavioral (Sterelny & Griffiths [1999], pp. 184–5; Sokal & Sneath [1963]). The central assumption behind a phenetic species concept is that any two members of a species will be more similar to each other than to any individual in another species. The hierarchical Linnaean system of classification that most biologists use today is largely the result of a phenetic approach to classification. Although the typical tree-like depiction of life *appears* to directly incorporate evolutionary histories, most of the contemporary use of the Linnaean system ultimately assumes that classification by *morphological* criteria will also separate species by evolutionary history. Biologists are aware, however, that selective evolutionary forces can cause different species with only distant common ancestry to converge upon similar morphologic features by *adapting* similarly to a shared or very similar design problem.

Phenetic species concepts are often rejected on theoretical grounds because there can be individuals that are less similar to conspecifics but should seemingly still belong to that species (Sterelny & Griffiths [1999], p. 184). Nonetheless, phenetic methods are commonly adopted by virologists attempting to classify viruses. Lawrence et al. ([2002]) lists some of the many ways virologists have attempted to classify viruses phenetically, including host range, pathogenic nature, overall morphology, and DNA sequence relatedness (Lawrence et al. [2002], p. 4892). The problem with these phenetic approaches is the relatively 'scattered' distribution that results; using one set of features for classification isolates organisms from other individuals that are otherwise very similar.

With the automated DNA sequencing technology presently available, one might expect that biologists could reach a definitive phenetic approach to virus classification using comparative genomics, that is, through a comparison of complete viral genomes. However, using complete genomes as the data for virus taxonomy presents biologists with a dilemma. Either one can use every base from complete genomes to make the comparison or one can privilege parts of the genome that are deemed more important, e.g., sequences that are more conserved among the genomes in question. On the first horn of the dilemma, because of prolific HGT, average similarity taken over many characteristics obscures the different histories of the different parts of mosaic viruses. Such an approach creates a taxonomy with little connection to the actual phylogeny and also little connection to the mosaic variation within viral genomes. The mosaicism that results from the prolific HGT often rules out the distinct clusters of bacteriophages in morphospace that pheneticists would call species. On the second horn of the dilemma, biologists take some phage genes as more important for the sake of classification (i.e., protein capsid genes, genes for penetration and/or extrusion, genes for viral genetic material replication, etc.). For example, Caroline Proux et al. ([2002], p. 6035) suggest, 'basing phage taxonomy on comparative genomics of a single structural gene module (head or tail genes).' However, such a distinction is arbitrary. By focusing only on a single structural gene module, one ignores the evolutionary history of non-structural modules, some of which have as much or more claim to be the basis for a correct classification. It is possible for two different bacteriophages to share head genes while the majority of the phages' remaining genes are radically different. For example, the *Shigella flexneri* phage SfV has HK97-like head genes and phage Mu-like tail genes (Allison et al. [2002]; see also Figure 1). Phenetic species concepts are ill-suited to bacteriophage taxonomy.

3.2 The biological species concept

Perhaps the most widely known species concept among biologists is Ernst Mayr's biological species concept (BSC). Mayr defines biological species as

'groups of interbreeding natural populations that are reproductively isolated from other such groups' (Mayr [1996], p. 264). Unlike pre-Darwinian essentialist approaches, the BSC allows for significant variation within a species while still providing a criterion (i.e., being a member of a particular reproductively isolated community) for membership in the species. The obvious lacuna concerns how to classify asexual organisms, a weakness Mayr himself acknowledges (Mayr [1996], p. 266; Ghiselin [2002], p. 157). Because viruses are asexual they do not form 'groups of interbreeding populations' in Mayr's sense, and are therefore not subject to classification under the BSC. To circumvent this problem, an advocate of the BSC might argue that there exists a viral analog to 'reproductive isolation' and thus create an extended BSC more applicable to viruses. For example, viruses reproduce via infection of a host cell, and thus host range might be such an analog. Unfortunately, the overlapping nature of viral host cell ranges means that there are no effective barriers to gene flow between very different viruses, and thus the extended BSC fails to properly classify viruses.

Lawrence et al. ([2002]) argue that because of the importance of HGT mechanisms in viruses, any virus classification system must take these mechanisms into consideration. Perhaps HGT mechanisms can act as *cohesion mechanisms*, much like sexual reproduction in higher organisms (Templeton [1989]; Sterelny & Griffiths [1999]; Lawrence et al. [2002]). For example, it is hypothesized that the high degree of conservation of gene order found across large groups of bacteriophages reflects an adaptation to increase the chances of successful HGT. In the Mu-like group of phages, the structural genes appear roughly in the same order in the various genomes (Morgan et al. [2002]; Figure 2). However, this approach also fails because, once again, the boundaries between purported species are not clear. Arranging viruses into classes because they are 'more likely' to swap genetic material with members of that group is an ineffective categorization due to the lack of discrete levels of propensity to swap genetic information. The situation is quite different from sexual species. Mayr avoids this problem because he assumes that there are discrete levels of propensity to swap genetic material in species that are classifiable by BSC. For him, there is a distinct difference between the propensity of genetic exchange among individuals that are members of a species and the propensity of genetic exchange among individuals from different species. Such an assumption cannot be justified for bacteriophages.

3.3 Phylogenetic species concepts

Phylogenetic species concepts define membership in a species not by synchronic characteristics, but by virtue of evolutionary and ancestral histories (Wiley [1978]; Mishler & Brandon [1987]). If the history of life is depicted as a

branching tree, a species (in the phylogenetic sense) is a lineage between two speciation events (de Queiroz [1999]; Mishler & Brandon [1987]). To be complete, a phylogenetic species concept requires an account of what constitutes a speciation event. It is one thing to say that a species consists of a particular lineage in the tree of life, but without defining what constitutes the beginning and endpoint of a lineage, there remains the possibility that a member of one lineage (at or near the point of 'speciation') might, by all measurable criteria, be indistinguishable from members of another lineage. Nonetheless, proponents of the phylogenetic species concept think that, at least in principle, an adequate account of when one species splits into new species can be developed.

HGT raises the opposite problem. Can two species come together to form a new species? For example, a 'daughter' phage may be produced from half of 'parent' phage X and half of 'parent' phage Y. Traditional species on the phylogenetic account have one unique history, although some have tried to accommodate hybrid plant species within orthodox species concepts. Wiley, for example, classifies a plant hybrid between genus A and genus B as a species in either genus (but not both), gives it a special designation, and thereby retains a Linnaean classification (Wiley [1981]). It is unclear to us whether such an approach is practical if every organism is multiply and thoroughly hybrid or mosaic as it is in bacteriophage evolution. His approach also introduces an arbitrary choice of genus when classifying a hybrid between two or more genera. Prolific HGT shows how many different parts of a viral genome can have many different histories. Unless the phylogenetic account is supplemented by an account of species fusion, there is no guarantee that the 'offspring' of a particular virus will be a member of only one lineage. This presents a definite problem for the phylogenetic approach, because a fundamental (but perhaps not essential) presupposition of orthodox taxonomy is that any particular organism is a member of one and only one species. We will return to this problem in Section 5.

3.4 The ecological species concept

The ecological species concept (ESC) determines species membership by the environmental niche occupied by a lineage of organisms (Van Valen [1976]). There are several problems with applying this concept in virology. First, it is difficult to define what an ecological niche *is* for a virus. One plausible candidate for the definition of a viral niche is the physiological states of the interior of an infected cell. However, which cellular conditions are relevant to describing a niche and which are not is a very difficult question to answer. Another candidate is the local environment that phages occupy between host infections. While we concede that these extracellular environments may sometimes play a selective role, we find that it is counter-intuitive to define an ecological niche as an environment in which the agent in question does little but sit or drift.

Further, given the abundance of bacteriophages, one would be challenged to find a terrestrial extracellular environment they do *not* occupy. The interior of an infectable bacterial cell seems the better choice for the ecological niche of a phage. The ESC then is prone to decay into merely using host range to classify bacteriophages. However, very different bacteriophages infect the same hosts and some bacteriophages can infect a *wide* variety of bacteria. Second, two arguably different species can converge on the same ecological niche without being in direct competition with one another. If biologists could identify lineages independently of niches, then perhaps the ESC could solve this problem. Nonetheless, even if biologists could define the niche for a particular bacteriophage, it is not clear how to individuate lineages in reticulated bacteriophage history. (We discuss this point further in Section 5). Finally, bracketing the above problems, the difficulty of discovering the host range of a phage in the wild, as opposed to the lab, imposes severe practical limitations on determining the niche of a given phage. The ESC is ultimately unsuccessful for classifying viruses.

3.5 Homeostatic property cluster species

Richard Boyd argues for a theory of natural kinds called homeostatic property clusters (HPC) that he later applied to biological species (Boyd [1991], [1999]). His position has been endorsed by some philosophers of biology (Griffiths [1999]; Wilson [1999]) and rejected by others (Ereshefsky [2001]; Ghiselin [2007]; Ereshefsky and Matthen [2005]). Boyd claims that there is a type of species essentialism and that biological species are paradigmatic cases of natural kinds. Abridging slightly, Boyd's idea of HPC has many parts:

1. There is a family (F) of properties that are contingently clustered in nature. . .they co-occur in an important number of cases.
2. Their co-occurrence is, at least typically, the result of. . .a sort of homeostasis.
3. The homeostatic clustering of the properties in F is causally important. . . important effects are produced by [their] conjoint occurrence . . .
4. There is a kind term, *t*, that is applied to things in which homeostatic clustering of most of the properties in F occurs.
5. *t* has no analytic definition; . . . all or part of the homeostatic cluster F, together with some or all of the mechanisms that underlie it, provide the natural definition of *t*.
6. Imperfect homeostasis is nomologically possible or actual. . .
7. Moreover, there will be many cases of extensional indeterminacy. . .

> 8. [There will be cases] such that no rational considerations dictate whether or not they are to be classed under *t*... (Boyd [1999], pp. 143–4).

Boyd's theory places importance on common synchronic properties, yet provides a way in which a biological agent may be a member of a species without containing some of these properties. He further allows that there are times when an individual is unable to be placed in any species category. He emphasizes the fact that the clustering of these properties is 'causally important', which is essentially a way of including a role for diachronic properties in his account. Nonetheless, we agree with Ereshefsky and Matthen ([2005]) that ultimately, Boyd's account uses similarity, not history, as 'the final arbitrator of species sameness.' (Ereshefsky and Matthen [2005], p. 17) In particular, Boyd needs an account of the individuation of clustering mechanisms and such an account would seem to rely on similarity rather than historical properties. On the other hand, HPC is not a pure phenetic account as it appeals to the similarity of the causes of clustering rather than the similarity within clusters.

When applied to biological species, Boyd's approach shares much with Templeton's cohesion species concept (CSC) (Templeton [1989]). As Templeton puts it, his approach, 'defines species in terms of the mechanisms yielding cohesion...' (Templeton [1989], p. 12). Templeton identifies two major kinds of cohesion mechanisms: those that influence genetic exchange or gene flow and those that influence genetic drift and natural selection. He does not consider prolific HGT, but only more conventional genetic exchange via sex. We think that HGT is a mechanism that influences gene flow much like sex does for sexual species and so should be considered a cohesion mechanism or at least part of one. He defines a species as, 'the most inclusive group of organisms having the potential for genetic and/or demographic exchangeability' (Templeton [1989], p. 25). The obvious difficulty in applying this definition to bacteriophages is that all phages have the potential for genetic exchangeability and thus one would be forced to say that practically all bacteriophages form one very large species. On the other hand, Templeton's discussion of cohesion mechanisms is more concrete than Boyd's notion of homeostasis and can be used to supplement the HPC view applied to biological species.

The HPC definition of species is more plausible for viruses than the concepts so far considered. It allows for differences among members of the same species, and no particular characteristic is itself a requirement for membership of that species. Nonetheless, bacteriophages are challenging for this species concept for reasons similar to those that challenge Templeton's CSC. As argued above, phages are so thoroughly mosaic that orthodox phenetic approaches to classification are ultimately unable to separate viruses into meaningful distinct groups of species satisfactorily while also respecting the diversity of their evolutionary histories. While Boyd's approach may cluster some bacteriophage properties,

due to the extensive mosaicism, many clusters of properties will overlap. Consider, for example, bacteriophage gene order. There are selection pressures regulating the relative order of the genes along bacteriophage genomes and properties that capture the rough gene order would be good candidates for being homeostatically clustered (Morgan et al. [2002]). One could identify a causally important cluster of properties, $P_1, \ldots, P_n$, that collectively capture a rough ordering of the genes in a population of viral genomes and that persist because of a 'homeostasis' driven by HGT. An otherwise functional rearrangement of a bacteriophage genome that loses a significant number of $P_1, \ldots, P_n$ would most likely result in a decrease in the long-term fitness of the genotype due to a lower probability of successful HGT. One of the surprising results of recent viral genomics is that very different phages can share a significant number of $P_1, \ldots, P_n$. These results suggest that Boyd's approach has the awkward consequence that phages that should intuitively be separated are, in fact, members of the same species.

Unfortunately for Boyd's view, the problem generalizes to other types of properties. HGT is arguably the most important clustering mechanism in bacteriophage evolution. Any phage can exchange genetic material with any other phage, either directly or through intermediaries. Because of the lack of boundaries of HGT within the entire bacteriophage population and the long history of exchanges among very different phages, biologists observe overlapping clusters of viral properties, whether they are properties of the respective genomes or properties of phenotypic expression (Hendrix et al. [1999]). Consequently, applying the HPC species concept to mosaic phages would lead to the classification of all phages as one (or only relatively few) species. In other words, prolific borderless HGT causes overlapping clusters of bacteriophage properties with no clear species boundaries. If the HPC proponent attempted to divide the bacteriophages into a large number of species, as many virologists would like to do, she would fail to find discrete clusterings or discrete clustering mechanisms and would be forced to conclude that phages are unclassifiable, or as Boyd would put it, that we have a case of 'extensional indeterminacy' (Boyd [1999], p. 144).

Given the prolific HGT among bacteriophages, one might wonder why the entire population does not converge on the most optimal design. Relatedly, what stops the bacteriophage population from blending into one uniform type?[1] In response to this concern, let us make two points. First, persistent polymorphism can exist within populations that undergo frequent genetic exchange. One can find many examples of persistent polymorphism within sexual species; an obvious case would be sexual dimorphism itself. Second, bacterial ecologists

[1] We thank Richard Lewontin, Hamish Spencer, and an anonymous reviewer for pushing us to consider this question.

speculate that the overall composition and diversity of the entire population of host bacteria changes significantly over time, leading to a constantly changing environment for the parasitic bacteriophages (Thingstad [2000]). The nature of the population dynamics of the host bacteria might explain much of the polymorphism found in bacteriophages. Nonetheless, we acknowledge more work needs to be done for a complete explanation of the high degree of polymorphism found in bacteriophages.

4 Viruses and Species Taxonomy

What, then, is a practicing virologist to do? In the early 1990s, the ICTV issued criteria for classifying viruses in terms of species, genera, families, and orders. For viral species, the ICTV endorsed the following definition: 'A virus species is a polythetic class of viruses that constitutes a replicating lineage and occupies a particular niche' (Pringle [1991], p. 303). The members of a polythetic class have a number of properties in common but no one property that all members share that defines the class. In effect, ICTV has adopted a blend of the phylogenetic and ecological species concepts. Note that for this definition to be satisfactory, they need to be able to adequately define *lineage* and *niche* for viruses. ICTV also defines a viral genus as simply 'a group of [viral] species sharing certain common characters.' The grouping guidelines for virus families and orders are the same as for virus genera, i.e., families are groups of genera that share certain common characteristics, etc. (van Regenmortel & Mahy [2004], p. 9) Of course, this vague definition raises the obvious question: which characters should a virologist use to identify and rank taxa and which should she disregard?

The principal problem here, we believe, is that the ICTV attaches too much importance to the traditional Linnaean *hierarchical* classification system and underestimates the importance of HGT among viruses. Bacteriophage HGT would seem to require a classification system based on reticulation, if biologists want to attempt to capture the degree to which the genomes, morphologies, and various other features of viruses are convoluted.

5 Reticular Phylogenies

Let us consider, then, if a system based on reticulation is appropriate for virus classification. There has been some theoretical work to make reticulate networks represent biological relationships. (See, for example, Doolittle [1999]; Sneath [2000]; Makarenkov et al. [2006], and the special section of *Journal of Classification*, vol. 17). The idea of reticulation is not new to biology. A type of reticulation occurs within a sexual species: after a crossing-over event in meiosis, a resultant chromosome derives partly from one parent and partly from the other. It also occurs between species when distinct species of plants

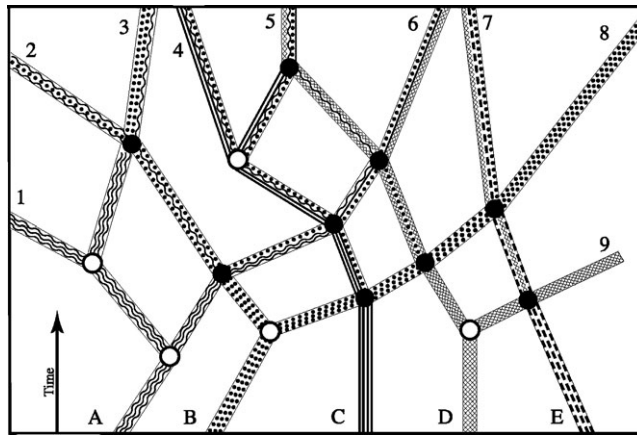

Figure 3. A representation of a reticulated phylogeny. Over time, phage lineages A–E undergo horizontal gene transfer (HGT) to create mosaic phage lineages 2–7. Lineage 9 goes extinct. Open circles represent splitting events and closed circles represent recombination events.

hybridize. Nonetheless, in bacteriophages the degree of reticulation is higher and is the rule rather than the exception. If the recombination boundaries occur only between genes, then one can represent the history of recombination as a set of trees, one for each gene. (If there has been no HGT, then ceteris paribus, all the gene trees will have the same topology). In other words, if mosaic bacteriophages consist of non-mosaic genes, then one can still represent the history of the genes that make up bacteriophages as sets of trees. However, if at least some of the genes that make up mosaic bacteriophages are also mosaics, which is likely the case, then reticulated representations can more accurately capture bacteriophage phylogeny.

One can represent the phylogenetic history of bacteriophages as a network or lattice (Figure 3). The straight patterned branches in Figure 3 then represent stretches of time during which the phage genomes remain unchanged. There are two types of vertices. First, there are more orthodox splitting events when one lineage splits into two or more daughter species. (We acknowledge that there are severe problems in defining what counts as a splitting of a viral lineage or even defining a viral lineage). Second, there are the vertices in which two species merge in a HGT event that creates at least one viable mosaic phage. The creation of a novel mosaic virus by HGT is represented by a significant change in the direction of a line through a closed circle. Notice that reticulated histories, unlike trees, need not have the topology required to divide organisms into nested monophyletic groups. Although there are independent problems with this view, such as how to deal with ancestral species, monophyly is a popular

necessary condition on definitions of species and higher taxa, especially among cladists. Biologists either have to seriously rethink what higher taxa are in cases of highly reticulated phylogenies or abandon the view that every biological agent fits into a hierarchical classification based on monophyly.

On a reticulated representation of viral history, a plausible suggestion is that a viral species is just a lineage between recombination and/or speciation events. This suggestion extends the phylogenetic species concept to account for reticulation. Although tempting, we think such a suggestion is too fine-grained. To stipulate that any viable phage that is created by a recombination event is a member of a new viral species is overly permissive. Just as biologists do not consider every mating and birth in the phylogeny of a sexual species, not all recombination events are equally important in the creation of a new species. Does the replacement of a single base pair constitute speciation? Probably not. Juhala et al. ([2000]) detail the exchange of small pieces of DNA called 'morons'. They consist of a transcriptional promoter, one or a small number of apparent genes, and a translational terminator. To gain or lose one of these pieces of DNA is probably not enough to create a new species. If it were enough, then every phage isolate characterized to date would be a new species. In fact, even the loss or gain of a number of functional genes might not be enough to create a new virus species. What criterion are we to use then to distinguish HGT that is significant enough to create a new species from that which is not? We can think of none without introducing an arbitrary threshold in what appears to be a continuum of severity of recombination. Further, the attempt to make such a demarcation amounts to an appeal to the phenetic method and brings with it the problems discussed above.

Even if we could solve the threshold problem, there are significant practical problems in determining an accurate highly reticulated bacteriophage phylogeny. Computationally, reconstructing reticulated networks (reticulograms) is more difficult than bifurcating trees. The best current algorithms for determining reticulated networks first construct a non-reticulated phylogeny (a tree) and then consider whether adding horizontal exchange improves the phylogeny. (See Makarenkov et al. [2006] for a review). In some cases, one starts with a tree of a gene thought not to have any horizontal exchange, e.g., a tree based on 16S rRNAs. One compares the topology of this tree with the topology of trees for genes in the same organisms that may have been involved in HGT. If there is a difference in topology, one considers whether adding horizontal exchanges to the second tree can account for the difference in topology. This type of algorithm works most effectively when there are relatively few horizontal exchanges and when you can identify a gene that is not involved in HGT. Neither of these conditions obtains for bacteriophage evolution. In fact, it is probable that in bacteriophage evolution horizontal exchanges are more

common than orthodox splitting events. Of course, it is possible that phylogenetic methods will be developed to solve this problem. All procedures, however, face the following issue: as we attempt to reconstruct more and more ancient recombination events, the unrecombined sections of DNA (sometimes called haplotypes) become smaller and smaller until they become too small to allow for phylogenetic reconstruction.

6 Conclusion

We have argued that the extensive mosaicism found in bacteriophages causes problems for any species concept, such as phenetic approaches, that uses synchronic properties to define a species. Bacteriophages' long history of HGT and the resulting reticulated phylogenies also create difficulties for diachronic historical approaches such as phylogenetic species concepts.

The pervasive mosaicism of phages and the inability to classify them effectively suggests an overall problem in defining what a biological species is. If we cannot devise an adequate definition of species that covers both viral *and* nonviral biological agents, then we find ourselves in a situation with the potential to erode parts of the traditional Linnaean hierarchy. (Ereshefsky [2001] expresses a related skepticism toward the usefulness of the Linnaean hierarchy). Consider bacteria, the 'vehicles' of the phages. Phages that enter the lysogenic life cycle have the ability to significantly alter the genome of their host. If prophage insertion leads to non-deleterious (perhaps even beneficial) effects, then at what point should biologists say a new species of bacteria has been created? The incoming genes need not be bacteriophage genes because some phages bring bacterial genes from a previous host with them. Bacteria, to varying degrees, *are* mosaics.

Multicellular life is also subject to HGT. Recent studies in virology have indicated that retroviruses have played a *significant* role in the evolutionary histories of other species (Van Blerkrom [2003]). The human genome provides a good example of the effect of viruses (particularly retroviruses) on nonbacterial species. Although there is some disagreement, the human genome is thought to be composed of at least one percent, but possibly as much as eight percent retroviral sequences (Sverdlov [2000]; Belshaw et al. [2004]; Parseval et al. [2003]; Khodosevich et al. [2002]). There is evidence that many retroviral sequences are an integral part of mammalian traits, e.g., the synthesis of immunosuppressive peptides during mammalian pregnancy that acts to deter the mother's immune system from conflicting with the fetus (Mallet et al. [2004]). Not only are viruses mosaics, but they also *cause* mosaicism in a wide variety of other organisms. We acknowledge that since the degree of mosaicism is lower in metazoans and that traditional species concepts apply much better than in the viral case, analogous arguments for abandoning the use of a species concept in multicelluar life are less compelling.

The Linnaean hierarchy is a vestige from an era in which prolific horizontal transfer and pervasive mosaicism were unimagined. Bacteriophages are evolving agents, but they do so without the confinement of species; they represent an incredible diversity of biological agents that have access to an enormous, common gene pool (Hendrix et al. [1999]). Each individual phage represents a miniscule sample from this gene pool. This biological arrangement presents a bold challenge to the traditional relations between natural selection and the origin of species. We should reconsider our model of how natural selection works. Natural selection does not necessarily act on populations to cause 'good and distinct' species to emerge. In bacteriophage evolution, adaptation to changing environmental conditions—mostly due to the population dynamics of the host bacteria—and prolific HGT results in individuals that are composed of convoluted chimerical genomes and phenotypes that reflect no one particular evolutionary history. There is good reason to think that the Earth's most abundant biological agent evolves without good and distinct species.

Acknowledgements

We thank Roger Hendrix, Jeffery Lawrence, Don Culberson, Lindley Darden, Todd Grantham, Stephanie Girard, Alan Love, Richard Lewontin, Hamish Spencer, Michael Ghiselin, E. O. Wiley, Stacey Welch, and an anonymous referee for their comments and help, December Williams for drawing Figure 3 and Katherine Cortez for drawing Figure 1. An audience at the ISHPSSB conference at Exeter in July 2007 also provided useful feedback. GJM thanks Spring Hill College for a Summer Research Grant.

Gregory J. Morgan
Department of Philosophy
Spring Hill College
4000 Dauphin St., Mobile
AL 36608, USA
gmorgan@shc.edu

W. Brad Pitts
College of Medicine
University of South Alabama
Mobile, AL 36688-0002, USA
wbp501@jaguar1.usouthal.edu

References

Allison, G. E., Angeles, D., Tran-Dinh, N. and Verma, N. K. [2002]: 'Complete Genomic Sequence of SfV, a Serotype-Converting Temperate Bacteriophage of *Shigella flexneri*', *Journal of Bacteriology*, **184**, pp. 1974–87.

Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paes, J., Burt, A. and Tristem, M. [2004]: 'Long-Term Reinfection of the Human Genome by Endogenous Retroviruses', *Proceedings of the National Academy of Sciences USA*, **101**, pp. 4894–9.

Boyd, R. [1991]: 'Realism, Antifoundationalism, and the Enthusiasm for Natural Kinds', *Philosophical Studies* **61**, pp. 127–48.

Boyd, R. [1999]: 'Homeostasis, Species, and Higher Taxa', in R. Wilson (*ed.*), *Species: New Interdisciplinary Essays*, Cambridge, MA: MIT Press, pp. 141–85.

Brüssow, H. and Desiere, F. [2006]: 'Evolution of Tailed Phages: Insights from Comparative Phage Genomics', in R. Calendar (*ed.*), *The Bacteriophages*, Oxford: Oxford University Press, pp. 26–36.

Eldredge, N. and Cracraft, J. [1980]: *Phylogenetic Patterns and the Evolutionary Process. Method and Theory in Comparative Biology*, New York: Columbia University Press.

Calendar, R. (*ed.*). [2006]: *The Bacteriophages*, Oxford: Oxford University Press.

Darwin, C. [1859]: *On the Origin of Species*, London: John Murray.

De Queiroz, K. [1999]: 'The General Lineage Concept of Species and the Defining Properties of the Species Category', in R. Wilson (*ed.*), *Species: New Interdisciplinary Essays*, Cambridge, MA: MIT Press, pp. 49–89.

Doolittle, W. F. [1999]: 'Phylogenetic Classification and the Universal Tree', *Science*, **284**, pp. 2124–8.

Dupré, J. [1993]: *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*, Cambridge, MA: Harvard University Press.

Ereshefsky, M. [1991]: 'Species, Higher Taxa, and the Units of Evolution', *Philosophy of Science*, **58**, pp. 84–101.

Ereshefsky, M. [2001]: *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*, Cambridge: Cambridge University Press.

Ereshefsky, M. and Matthen, M. [2005]: 'Taxonomy, Polymorphism, and History: An Introduction to Population Structure Theory', *Philosophy of Science*, **72**, pp. 1–21.

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. and Ball, L. A. [2005]: *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses*, San Diego, CA: Elsevier Academic Press.

Ghiselin, M. [1974]: 'A Radical Solution to the Species Problem', *Systematic Zoology*, **23**, pp. 536–44.

Ghiselin, M. [1987]: 'Species Concepts, Individuality, and Objectivity', *Biology and Philosophy*, **2**, pp. 127–43.

Ghiselin, M. [2000]: 'Species Concepts: The Basis for Controversy and Reconciliation', *Fish and Fisheries*, **3**, pp. 151–60.

Ghiselin, M. [2007]: 'Is the Pope a Catholic?', *Biology and Philosophy*, **22**, pp. 283–91.

Griffiths, P. [1999]: 'Squaring the Circle: Natural Kinds with Historical Essences', in R. Wilson (*ed.*), *Species: New Interdisciplinary Essays*, Cambridge, MA: MIT Press, pp. 209–28.

Hendrix, R. W. [2005]: 'Bacteriophage HK97: Assembly of the Capsid and Evolutionary Connections', *Advances in Virus Research*, **64**, pp. 1–14.

Hendrix, R. W., Lawrence, J. G., Hatfull, G. F. and Casjens, S. [2000]: 'The Origins and Ongoing Evolution of Viruses', *Trends in Microbiology*, **8**, pp. 504–8

Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. and Hatfull, G. F. [1999]: 'Evolutionary Relationships among Diverse Bacteriophages and Prophages: All the World's a Phage', *Proceedings of the National Academy of Sciences USA*, **96**, pp. 2192–7.

Hull, D. [1978]: 'A Matter of Individuality', *Philosophy of Science*, **45**, pp. 335–60.

Hull, D. [1987]: 'Genealogical Actors in Ecological Roles', *Biology and Philosophy*, **2**, pp. 168–83.

Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F. and Hendrix, R. W. [2000]: 'Genomic Sequences of Bacteriophages HK97 and HK022: Pervasive Genetic Mosaicism in the Lambdoid Bacteriophages', *Journal of Molecular Biology*, **299**, pp. 27–51.

Khodosevich, K., Lebedev, Y. and Sverdlov, E. [2002]: 'Endogenous Retroviruses and Human Evolution', *Comparative and Functional Genomics*, **3**, pp. 494–8.

Kitcher, P. [1984]: 'Species', *Philosophy of Science*, **51**, pp. 308–33.

Lawrence, J. G., Hatfull, G. F. and Hendrix, R. W. [2002]: 'Imbroglios of Viral Taxonomy: Genetic Exchange and Failings of Phenetic Approaches', *Journal of Bacteriology*, **184**, pp. 4891–905.

Makarenkov, V., Kevorkov, D. and Legendre, P. [2006]: 'Phylogenetic Network Reconstruction Approaches', Special issue, Applied Mycology and Biotechnology, International Elsevier Series, vol. 6, *Bioinformatics*, pp. 61–97.

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G, Duret, L. and 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**, pp. 1731–6.

Mayr, E. [1996]: 'What Is a Species and What Is Not?', *Philosophy of Science*, **63**, pp. 262–77.

Mishler, B. and Brandon, R. [1987]: 'Individuality, Pluralism, and the Phylogenetic Species Concept', *Biology and Philosophy*, **2**, pp. 397–414.

Morgan, G. J., Hatfull, G. F., Casjens, S. and Hendrix, R. W. [2002]: 'Bacteriophage Mu Genome Sequence: Analysis and Comparison with Mu-like Prophages in Haemophilus, Neisseria and Deinococcus', *Journal of Molecular Biology*, **317**, pp. 337–59.

Parseval, N., Lazar, V., Casella, J., Benit, L. and Heidmann, T. [2003]: 'Survey of Human Genes of Retroviral Origin: Identification and Transcriptome of the Genes with Coding Capacity for Complete Envelope Proteins', *Journal of. Virology*, **77**, pp. 10414–22.

Pringle, C. R. [1991]: 'The 20th Meeting of the Executive Committee of the International Committee on Virus Taxonomy', *Archives of Virology*, **119**, pp. 303–4.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G. F., Desiere, F. and Brüssow, H. [2002]: 'The Dilemma of Phage Taxonomy Illustrated by Comparative Genomics of Sfi21-like *Siphoviridae* in Lactic Acid Bacteria', *Journal of Bacteriology*, **184**, pp. 6026–36.

Sneath, P. H. A. [2000]: 'Reticulate Evolution in Bacteria and Other Organisms: How Can We Study It?', *Journal of Classification*, **17**, pp. 159–63.

Sober, E. [1984]: 'Sets, Species, and Natural Kinds: A Reply to Philip Kitcher's "Species"', *Philosophy of Science*, **51**, pp. 334–41.

Sokal, R. and Sneath, P. H. A. [1963]: *Principles of Numerical Taxonomy*, San Francisco: W. H. Freeman.

Sterelny, K. and Griffiths, P. [1999]: *Sex and Death: An Introduction to Philosophy of Biology*, Chicago: Chicago University Press.

Sverdlov, E. [2000]: 'Retroviruses and Primate Evolution', *BioEssays*, **22**, pp. 161–71.

Templeton, A. R. [1989]: 'The Meaning of Species and Speciation: A Genetic Perspective', in D. Otte and J. A. Endler (*eds*), *Speciation and Its Consequences*, Sunderland, MA: Sinauer, pp. 3–27.

Thingstad, T. F. [2000]: 'Elements of a Theory for the Mechanisms Controlling Abundance, Diversity, and Biogeochemical Role of Lytic Bacterial Viruses in Aquatic Ecosystems', *Limnology and Oceanography*, **45**, pp. 1320–8.

Van Blerkrom, L. [2003]: 'Role of Viruses in Human Evolution', *Yearbook of Physical Anthropology*, **46**, pp. 14–46.

Van Valen, L. [1976]: 'Ecological Species, Multispecies, and Oaks', *Taxon*, **25**, pp. 223–39.

Van Regenmortel, M. H. V. and Mahy, B. W. J. [2004]: 'Emerging Issues in Virus Taxonomy', *Emerging Infectious Diseases*, **10**, pp. 8–13.

Wiley, E. O. [1978]: 'The Evolutionary Species Concept Reconsidered', *Systematic Zoology*, **27**, pp. 17–26.

Wiley, E. O. [1981]: *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*, New York: John Wiley.

Wilson, R. A. [1999]: 'Realism, Essence, and Kind: Resuscitating Species Essentialism?', in R. Wilson (*ed.*), *Species: New Interdisciplinary Essays*, Cambridge, MA: MIT Press, pp. 187–207.