Protein & Cell

Protein & Cell

REVIEW

Romance of the three domains: how cladistics transformed the classification of cellular organisms

Chi-Chun Ho[1], Susanna K. P. Lau[1,2,3,4✉], Patrick C. Y. Woo[1,2,3,4✉]

[1] Department of Microbiology, The University of Hong Kong, Hong Kong, China
[2] State Key Laboratory of Emerging Infectious Diseases, The University of Hong Kong, Hong Kong, China
[3] Research Centre of Infection and Immunology, The University of Hong Kong, Hong Kong, China
[4] Carol Yu Centre of Infection, The University of Hong Kong, Hong Kong, China
✉ Correspondence: pcywoo@hkucc.hku.hk (P. C. Y. Woo), skplau@hkucc.hku.hk (S. K. P. Lau)
Received June 21, 2013   Accepted July 1, 2013

ABSTRACT

Cladistics is a biological philosophy that uses genealogical relationship among species and an inferred sequence of divergence as the basis of classification. This review critically surveys the chronological development of biological classification from Aristotle through our post-genomic era with a central focus on cladistics. In 1957, Julian Huxley coined cladogenesis to denote splitting from subspeciation. In 1960, the English translation of Willi Hennig's 1950 work, *Systematic Phylogenetics*, was published, which received strong opposition from pheneticists, such as numerical taxonomists Peter Sneath and Robert Sokal, and evolutionary taxonomist, Ernst Mayr, and sparked acrimonious debates in 1960–1980. In 1977–1990, Carl Woese pioneered in using small subunit rRNA gene sequences to delimitate the three domains of cellular life and established major prokaryotic phyla. Cladistics has since dominated taxonomy. Despite being compatible with modern microbiological observations, i.e. organisms with unusual phenotypes, restricted expression of characteristics and occasionally being uncultivable, increasing recognition of pervasiveness and abundance of horizontal gene transfer has challenged relevance and validity of cladistics. The mosaic nature of eukaryotic and prokaryotic genomes was also gradually discovered. In the mid-2000s, high-throughput and whole-genome sequencing became routine and complex geneologies of organisms have led to the proposal of a reticulated web of life. While genomics only indirectly leads to understanding of functional adaptations to ecological niches, computational modeling of entire organisms is underway and the gap between genomics and phenetics may soon be bridged. Controversies are not expected to settle as taxonomic classifications shall remain subjective to serve the human scientist, not the classified.

INTRODUCTION

"The empire, long divided, must unite; long united, must divide."—*The Romance of the Three Kingdoms*, Guanzhong Luo

The history of biological classification is long and tortuous (Fig. 1). As biological beings are, obviously, the earliest things that our ancient ancestors encountered and classification allows entities to be better and more easily understood, the long history of biological classification should not seem surprising. In terms of lasting influence, Aristotle's (384 BC–322 BC) classification of living organisms can be ranked among that of Carl Linnaeus (1707–1778) and modern taxonomists such as Ernst Mayr (1904–2005), Willi Hennig (1913–1976), Peter Sneath (1923–2011), Robert Sokal (1926–2012), Carl Woese (1928–2012) and Thomas Cavalier-Smith (1942–).

The Greek philosopher Aristotle proposed a ladder of life, i.e. the *scala naturae*, to classify and rank living as well as non-living entities. According to the ladder, humans are ranked under God and the angels but distinctly above all other animals and plants. Animals were ranked by their perceived level of physical strength, intelligence, sensitivity and even habitat, such that
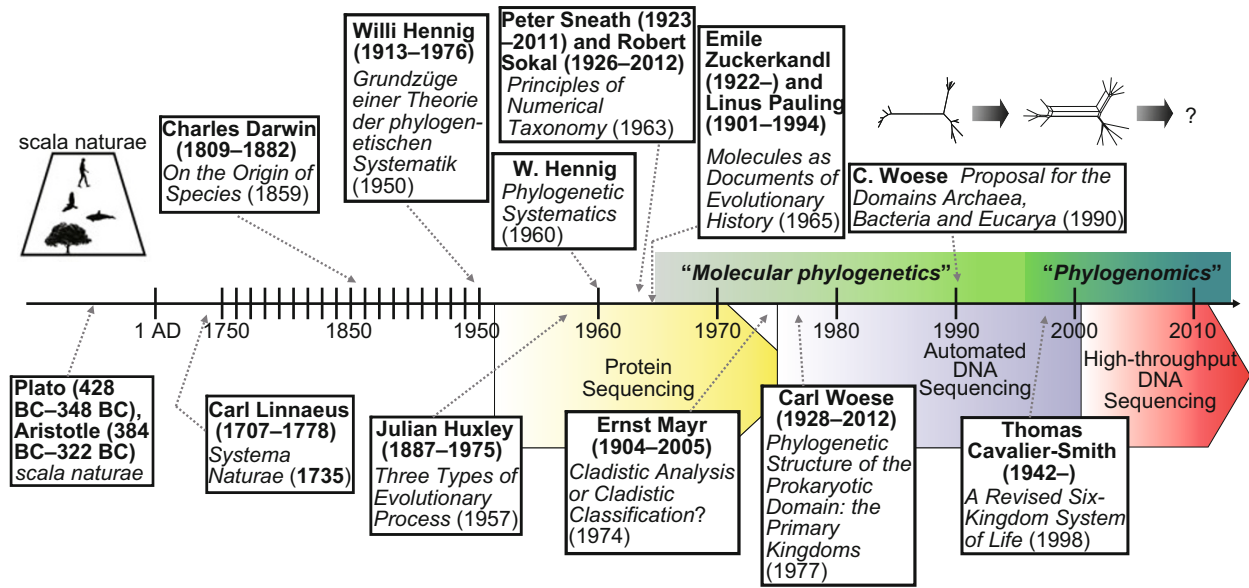
Figure 1. Time line of events. The cladistic transformation of biological classification. Horizontal axis denotes the time sequence of important ideologies or publications. The introduction of molecular techniques and the respective disciplines are shown towards the right as shaded boxes.

beasts such as lions and elephants were ranked among the highest animals and birds, capable of flight in air, were ranked higher than fishes. Plants were considered lacking intelligence and sensitivity and therefore lower than the animals. It should be noted that the classification was not strictly biological, as the lower levels of the scale actually went on to include the minerals such as diamond (with hardness and luster, the highest mineral), marble and grit (the lowest mineral); the *scala naturae* portrayed the then prevalent belief that the biological beings were integral to the world created by God and entities therefore had a logical hierarchy or order which was based on reason and logic. Even if we do not consider the creationist perspective intrinsic to this early classification system, the hierarchical view of life forms can also be found to influence such work as Linnaeus's *Systema Naturae* and the rampant notion that certain extant organisms are higher (e.g. *Homo sapiens*) and some are more primitive (e.g. *Escherichia coli*). We note, from an evolutionary point of view, such comments are not justified as both humans and bacteria have the same origin from the last universal common ancestor and have underwent evolution and natural selection for the same length of time. Indeed, if the actual amount of sequence evolution is taken into account, it becomes even harder to justify the opinion that humans are more evolved than even monkeys and apes (Li and Tanimura, 1987).

This review has two objectives: to present a critical appraisal of the history and development of cladistics and to evaluate its impact on the classification of cellular organisms from our understanding as microbiologists. Starting with such pioneers as Julian Huxley (1887–1975), who formally proposed the term *cladogenesis*, and Willi Hennig, the German biologist who is widely honoured as the "Father of Cladistics", we examine the background and birth of cladistics as well as some of the early debates. Then, as cladistics impacted biological classification, we revisit the views of some of the opinion leaders, such as Ernst Mayr, who proposed the famous biological species concept and advocated a genomic approach to evolutionary biology yet disagreed with the cladistic classification by Hennig, and Carl Woese, who brought cladistics to unprecedented heights by his phylogenetic-based proposal of the three-domain classification, i.e. *Eukarya*, *Bacteria* and *Archaea*. We also examine the disagreements from Thomas Cavalier-Smith and Stephen Jay Gould to this widely-accepted paradigm. Finally, given the increasing recognition of the nature and extent of horizontal gene transfer as more than sporadic occurrences which may transcend cytologic and phylogenetic barriers of the kingdoms and domains, we reassess and try to justify the utility and relevance of cladistics in this genomic era: still divided by clades, or are we back to formlessness, darkness and void?

HISTORICAL OVERVIEW OF CLADISTICS: THE REBELLION

On 7th September 1957, Julian Huxley's two-page article, titled *The Three Types of Evolutionary Process*, was published (Huxley, 1957). He extended the proposal by Bernhard Rensch (1954) and defined the term cladogenesis "to denote all splitting, from subspeciation through adaptive radiation to the divergence of phyla and kingdoms". Although some have suggested that terminology of a clade had been used earlier in the literature by Lucien Cuénot (1866–1951) or Ernst Haeckel (1834–1919), it has been pointed out that their use of the

term merely as a category in classification was different from Huxley's (Williams and Ebach, 2009) and Huxley was the first to put the concept in an evolutionary perspective and defined clades as "delimitable monophyletic units". This definition, however, represented a mix of subjective and objective criteria. While monophyletic units can be pin-pointed exactly in a strictly bifurcating phylogeny, biologists occasionally disagree as to whether a group of organisms is delimitable from another. The delimitability was later dropped from the definition when Huxley presented at the Systematics Association (Huxley, 1959). In modern molecular phylogeny, the clade as a monophyletic group is defined as a group of organisms that have been derived from a common ancestor, including the common ancestor itself (Page and Holmes, 1998; Graur and Li, 2000). Cladogenesis is now also known as a synonym for genetic speciation (US National Library of Medicine, 2006).

It is unknown whether Huxley anticipated the rather acrimonious debate between the cladists and pheneticists in the decades following his initial proposal. Believing that convergence and parallelism were difficult to detect given the incompleteness of the fossil record and "a thoroughgoing phylogenetic classification does not recognize such (non-monophyletic) grades at all", Huxley suggested a double system to express both biological improvement and evolutionary relationships (Huxley, 1959). This ambivalence was not accorded by contemporary cladists. When Hennig later introduced his landmark work, *Phylogenetic Systematics*, to the English world, he noted that "There is, in fact, a widespread notion that phylogenetic systematics, at least in those groups of animals for which no fossil finds are available, possesses no method of its own… This notion is false" (Hennig, 1965). Notably, Hennig's original publication of the *Grundzüge einer Theorie der phylogenetischen Systematik* (in English "Outlines of a Theory of Phylogenetic Systematics") actually predates that of Huxley (Hennig, 1950). In the later English translation of the work, known also as *Phylogenetic Systematics*, Hennig established the major cladistic principles, including the genealogical relationship among species as the basis of clades; shared-derived (synapomorphic) characters as the sole source of phylogenetic evidence; and the consistency of a proposed phylogeny with multiple lines of evidence (Hennig, 1966). While not discrediting the fossil record, Hennig logically argued, from his emphasis on synapomorphic traits, that the shared-ancestral (plesiomorphic) characters that had been the focus of fossil analysis were relatively unimportant for establishing a phylogeny; from a phylogenetic standpoint, he asserted that, instead, apomorphic characters on fossils should be sought and analysed after careful exclusion of convergence and retrogression (Hennig, 1966).

As a side note, Hennig was not the first to point out the difference between homology and analogy. As early as the late 19th century, Darwin has rectified the definition of homology to include common ancestry in addition to mere structural similarities in *The Origin of Species* (Darwin, 1859), and the concept was further extended by Hubbs (1944) to encompass functional similarities from common descent. In the meantime, some however insisted that homology should be limited to the description of homologous structural similarities instead of functional, physiological and behavioral correspondence from common descent (Boyden, 1947). In retrospect, this reluctance to change was understandable because of the difficulty if not a practical impossibility in exactly determining the function of a structure even given a comprehensive and well-preserved fossil record. The use of molecular techniques to determine the homology and function of biological structures was not foreseen.

Hennig's cladistics was not universally accepted in the systematics community. As the notion *Phylogenetic Systematics* implies, Hennig's method had the goal of changing the grouping and ranking of organisms based on cladistic principles, i.e. by the recency of common descent. While some zoologists voiced their support for the new method: "For if we do desert it, systematics inevitably returns to its pre-Darwinian status as a method of merely pigeonholing information." (Myers, 1952); or accepting and promoting it through field studies (Brundin, 1966), prominent taxonomists such as Ernst Mayr and Robert Sokal strongly rejected the use of cladistics in classification. Mayr believed that the branching pattern of a cladogram did not convey as much information as an "evolutionary classification" which took into account the adaptive features of different taxa and frequency of branching, and that the cladistic proposal, once implemented, would "produce classifications that are unbalanced and meaningless" (Mayr, 1974). Evolutionary systematists, represented by the Simpson-Mayr school and supporters, recognized cladistics as an important contribution to systematics but considered the application of cladistic principles to re-classify living organisms as being "neither necessary nor desirable" (Ashlock, 1974).

As a generation of microbiologists nurtured in modern phylogenetic methods, we note that many of the polarized arguments between the more phenetically-inclined "evolutionary taxonomists" and cladists could have been harmonized if they had been able to obtain the wealth of genetic information now available. Arguments stemming from the impracticability of cladistics, such as the difficulty of determining homology and the direction of evolutionary relationships; differentiation between parallel and convergent evolution; and doubts about the information content of ancestral characters (Mayr, 1974), have gradually been rendered obsolete as molecular techniques, such as serum protein electrophoresis (Johnson and Wicks, 1959), multi-locus enzyme electrophoresis (Avise, 1974), DNA-DNA hybridization (McCarthy and Bolton, 1963; Hyer et al., 1964), protein sequence comparison (Margoliash et al., 1961; Zuckerkandl and Pauling, 1965) and, finally, DNA sequencing (Sanger et al., 1973, 1977), were being continuously developed and popularized. We, however, agree with Mayr's view that, as an inevitable nature, cladistics does not entirely reflect functional evolution, which is occasionally biologically significant; cladistic classifications therefore may not be good heuristics to phenotypic characters; and that the indiscriminate use of cladistics in taxonomy could lead to the generation of

an unmanageable number of extant and extinct taxa (Mayr, 1981). In these aspects, Sneath and Sokal's numerical taxonomy (Sneath and Sokal, 1962) and its underlying philosophy are of particular interest as they represent a methodological treatment of phenetics with practical value in classification and prediction. We shall return to this topic as we focus on bacterial identification and classification.

BIOLOGICAL CLASSIFICATION AFTER CARL WOESE: THE REIGN

The modern, Linnaean classification of biological organisms can be traced back to the 18th century, when the goal of classification was to facilitate the field work of naturalist-botanist-physician Carolus Linnaeus (1751). Presumably, the inference of a phylogenetic relationship was unimportant, given the once prevalent biblical belief that all species had originated from a special creation and had not been evolving (1611). The goal of a practical classification system, therefore, is to provide positive identification (What is this plant/animal/fungus?), delineation (What are the differences between these two groups of plants/animals/fungi?) and circumscription of organism groups (Do these plants/animals/fungi belong to the same species/genus/family?); cladistic classifications, to this end, have limited "practical value". Certain practical considerations from Linnaeus's days have persisted and remained central to branches of botany, zoology, as well as basic and medical microbiology. A phenetic approach is useful when the goal of classification is to assign static groups for a fixed number of extant species. Even though practicality is central to such and related approaches, they always necessitate subjective judgement (e.g. deciding if a phenotypic character is a biological improvement or difference worthy of taxonomic demarcation). However, if the aim of classification is beyond that of extant, observed biodiversity and the objective of inference includes past biological forms and functions, only a cladistic classification can provide useful information. In a mixed classification approach, such as the evolutionary classification by Mayr (1974, 1981), or the polyphasic taxonomy commonly employed in basic and clinical microbiology (Colwell, 1970), the cladistic component may actually facilitate the documentation of novel biodiversity.

In 1977, Carl Woese and George Fox published their phylogenetic analysis of some eukaryotic organisms (e.g. baker's yeast and common duckweed), their organelles (e.g. duckweed chloroplast) and some prokaryotic organisms, including a few methanogens (Woese and Fox, 1977). In their landmark study, instead of the two groups that would have been predicted by the phenetic, prokaryotic-eukaryotic delimitation, the organisms (and organelles) could be divided into three groups based on comparative cataloguing of nucleotide hexamers of their small subunit (SSU) ribosomal RNA (Balch et al., 1977; Fox et al., 1977); the prokaryotic organisms could be further divided into two groups, namely, (i) the "eubacteria" which included all the bacterial species considered typical in the study (e.g. *Escherichia coli* and *Bacillus firmus*) and also the orga-

nelle of an eukaryote (i.e. duckweed chloroplast), and (ii) the methanogens (e.g. *Methanobacterium thermoautotrophicum* and *Methanosarcina barkeri*), which Woese and Fox named "archaebacteria" (Woese et al., 1978). To justify their classification, which was mainly based on a rudimentary measure of ribosomal RNA sequence similarity, it was further pointed out that the cell walls of the methanogens they studied differed from that of other bacteria by lacking peptidoglycan, and the biochemical and transfer RNA modification pathways of the methanogens were distinct from those of other bacteria as well as the eukaryotic organisms (Woese and Fox, 1977).

The presumed monophyly of the prokaryotes was formally challenged thirteen years later, when the proposal for the domains *Archaea*, *Bacteria* and *Eucarya* was published (Woese et al., 1990a). By comparing the primary sequences and the inferred secondary structures of the SSU ribosomal RNA gene of the three groups; usage pattern of RNA polymerases; some limited bacterial fossil record; and a phylogenetic rooting strategy using a gene duplication event inferred to be ancestral to the divergence of the groups, a "natural" (i.e. cladistic) classification comprising three domains was formulated. In the rooted phylogenetic tree presented in the article, Domain *Archaea* was more closely related to Domain *Eucarya*, and two Kingdoms, Kingdom *Euryarchaeota* and Kingdom *Crenarchaeota*, were also proposed to contain two monophyletic lineages within Domain *Archaea*.

In the interim between the two proposals, Woese and his co-workers' numerous publications on prokaryotic phylogeny amounted to what we would call the "cladist's reign". Earlier, 16S rRNA fingerprinting analysis had supported the prokaryotic origin of the chloroplast (Zablen et al., 1975); 16S rRNA gene sequence analysis in this period supported the prokaryotic origin of the mitochondrion (Yang et al., 1985b). A number of prokaryotic species were classified or re-classified as *Archaea* based on 16S rRNA gene sequence evidence (Magrum et al., 1978; Woese et al., 1980b; Gupta et al., 1983; Woese et al., 1984a; Olsen et al., 1985; Achenbach-Richter et al., 1988; Burggraf et al., 1990). A methanogenic origin of cellular life was proposed, supported by the presumed rooting of the Tree of Life near the *Bacteria-Archaea* split (Woese, 1979). The basic framework on which detailed analysis of 16S rRNA gene sequences is based was established (Woese et al., 1980a; Noller and Woese, 1981). Traditional phenetic classification of cell wall deficient bacteria was revised (Woese et al., 1980b). More importantly, the major bacterial and archaeal phyla were established in this period based on rRNA gene sequence evidence (Paster et al., 1984; Woese et al., 1984b; Oyaizu and Woese, 1985; Paster et al., 1985; Woese et al., 1985; Yang et al., 1985a; Oyaizu et al., 1987; Weisburg et al., 1989; Yang and Woese, 1989; Woese et al., 1990, 1990c, 1990d). The cladistic-driven approach adopted by Woese ironically put Mayr's comment in perspective "this approach by trial and error… led to frequent changes in classifications" (Mayr, 1974), although, as we contend, Woese's cladistic analysis and re-classification were more stable and reliable than Mayr had expected. The

Protein & Cell

Protein & Cell

current taxonomic framework of prokaryotic systematics is still largely compatible with what has been established by Woese (Ludwig and Klenk, 2005).

The impact of Woese's cladistic classification approach was profound. As PCR and DNA sequencing became common molecular techniques by the early 1990s, a number of conserved primers capable of producing nearly full-length 16S rRNA gene amplicons for subsequent sequencing and analysis were being published for both prokaryotes (Weisburg et al., 1991) and eukaryotes (Medlin et al., 1988). Distant phylogenetic relationships, arguably beyond the reach of most phenetic methods, were elucidated among animal phyla (Field et al., 1988). Large-scale projects, such as the ARB project (Ludwig et al., 2004), Ribosomal Database Project (Olsen et al., 1992) and the Tree of Life Web Project (http://tolweb.org), were initialized in this period to help deduce the phylogenetic relationship among living organisms. Limited by cost and the early technical difficulties, the predominant target used for molecular phylogeny was the small subunit (SSU) ribosomal RNA gene, i.e. the 16S rRNA gene in prokaryotes and the 18S rRNA gene in eukaryotes, as its moderate length allowed relatively efficient amplification and sequencing for even large-scale studies and the presence of conserved and variable regions along this functionally conserved molecule allowed phylogenetic inference to be made at various depths; this legacy is self-evident by the sheer number of SSU rRNA gene sequences in various electronic databases (DeSantis et al., 2006; Pruesse et al., 2007; Cole et al., 2009; Woo et al., 2011).

As microbiologists, we think we are justified to mention a few facts occasionally neglected by early taxonomists, who were usually botanists or zoologists, although certain salient points, such as the classification of the larval and adult form of the same insect and quantitative variation of a character, have been properly addressed as the allomorphism of species (Hennig, 1950). These observations originate from our practice and experience in clinical and basic microbiology and are hereby generalized for the wider audience; the reader is invited to judge the importance of these observations in their own respective fields.

Organisms can have unusual phenotypic profiles. As phenetic techniques advanced, the characterization of microorganisms progressed from size, shape and staining characteristics to usually a battery of biochemical tests (Woo et al., 2000). We reckon that there has been a similar trend in the botanical and zoological sciences. Given biological variation, it is plausible that organisms with atypical phenotypic profiles compared with previously observed organisms used to establish the phenetic classification and/or identification scheme be ambiguously (Woo et al., 2001a; Lau et al., 2005, 2006b) or incorrectly identified in the phylogenetic or cladistic sense, even if state-of-the-art techniques, such as matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS), are used (Chan et al., 2012). Widely present, housekeeping genes (Woo et al., 2003a, 2003b; Lau et al., 2006a, 2011a) or clade-specific, synapomorphic genomic features (Ho et al.,

2011, 2012) can provide a relatively reliable framework on which classification can be based, as exemplified by the role of the 16S rRNA gene phylogeny in the prokaryotic taxonomic framework (Ludwig and Klenk, 2005).

Organisms may not live to manifest all phenotypic characteristics. It is well-known in clinical microbiology that certain important pathogens, such as *Mycobacterium tuberculosis*, the causative agent of tuberculosis, takes a relatively long time (6–8 weeks) to grow and manifest the subtlest phenotypic characteristics such as its volatile fatty acid profile (Woo et al., 2008). This length of time is, ostensibly, short compared with the animal and plant species with generation time exceeding tens of years and should not impress the zoologists and botanists. Yet it is only common knowledge that phenotypic characteristics of plants and animals, such as the presence and color of flowers, presence of appendages or cranial size, can be affected by environmental factors, the exact stage of growth and life cycle. A relevant example may be cited from our field: certain fungal species, collectively known as the dimorphic fungi, exhibit a yeast phase and a mould phase. Not only are the two phases very distinct by visual inspection, they are usually associated with very different profiles of metabolite production (Woo et al., 2010, 2012) and potential reproductive strategy (Woo et al., 2006). What may complicate the issue further is the reversible nature of the phase switch (Woo et al., 2003c); so if the switch had not been recognized and the two phases classified into relatively unrelated groups by a phenetic approach it could produce a taxonomy that is more perplexing than useful. It takes little imagination to realize the vaguely analogous situations of metamorphosis in some animal species and alternation of generations in certain plants, or perhaps the multi-staged life cycles of many protoctista (Burkholder and Glasgow, 1997).

Organisms may leave only molecular evidence for their previous presence. In clinical microbiology, due to various reasons, the infectious agent may not be recovered from samples taken from the patient (Woo et al., 2001c). Instead, an infective episode may only be demonstrated by means of immunological assays targeting part of the agent or host responses (Woo et al., 2001b). More recently, the advent of nucleic amplification techniques has allowed the direct detection of nucleic acids from clinical specimens to evidence presence of an infectious organism. The documentation of novel biodiversity by metagenomics (Tse et al., 2012) and complete genome sequencing for uncultured organisms (Hallam et al., 2006; Hongoh et al., 2008a, 2008b) therefore seem to be a natural extension to such. If one insists on characterizing a complete specimen of the organism before establishing a new taxon, he volitionally neglects the logical sequence evidence and implicitly attributes this abundance of biological sequences to non-evolutionary processes.

As microbiologists at the interface of clinical and basic microbiology, we are perceptive of the influence of numerical taxonomy and molecular cladistics on classification. It is worth pointing out that classification, as a taxonomic endeavor, is

a distinct step that is thought to precede the corresponding identification scheme for practical procedures (Brenner et al., 2005). Therefore, the modern use of "numerical" or "phenetic" identification techniques, such as similarity indices from biochemical profiling, DNA-DNA hybridization and BLAST comparison of gene sequences, do not represent a breakdown of the current, cladistic-based taxonomy; this is because the goal of clinical or field identification is often the binning of an organism to an established taxon and only occasionally the recognition of novel species; a phylogeny is not always useful to identification. To establish the definitive identity of an organism for the basic and evolutionary researches, however, the rigorous construction of a phylogenetic hypothesis is usually required. As remarked by Page and Homles, the maximally informative or predictive classifications advocated by pheneticists and evolutionary taxonomists "may or may not be actual phylogenies; if they are, this is merely a happy accident" (Page and Holmes, 1998).

The impact of numerical taxonomy, in a sense the *holy grail* of phenetics, on classification is hereby examined. We feel that we are from the right field to comment on such since in reviewing the three decades of development of numerical taxonomy, Sneath particularly noted its success in microbiology and cited the proportion of publications in the *International Journal of Systematic Bacteriology* including "numerical relationships" as support (Sneath, 1995). Whilst it is apparent that common identification techniques in clinical microbiology, including the analytical profile index (API) and related methods (Titsworth et al., 1969; Washington et al., 1971), work by the numerical coding of arrays of biochemical test results with subsequent manual or computerized search in databases to achieve a "numerical identification", whether this also represents "numerical taxonomy" is disputable. Numerical taxonomy, as established by Sneath and Sokal in the early 1960s (Sneath and Sokal, 1962; Sokal, 1963), is a proposition against taxonomic groupings based on phylogenetic deductions. In their 1962 article, Sneath and Sokal argued for the separation of the taxonomic process from phylogenetic speculations as they believed that phylogenetic classifications have logical pitfalls: by providing just phylogenetic hierarchies representing the time sequence of divergence, a phylogenetic classification does not contain the maximum possible information for the researcher as would be achieved by a numerical taxonomy using cluster analysis (Sneath and Sokal, 1962). With such, Sneath and Sokal hoped to achieve "stable phenetic groups" in taxonomy which would also help find "features best suited to making a diagnostic key" (Sneath and Sokal, 1962). Later, they also collaborated with other groups to promote the numerical taxonomy method to the classification of autoimmune diseases (Jones et al., 1970, 1973) and medically important genera of bacteria (Broom and Sneath, 1981; Sneath et al., 1981; Bridge and Sneath, 1983). In 1970, Colwell published his application of numerical taxonomical techniques to derive a polyphasic taxonomy for the *Vibrio* genus (Colwell, 1970). Still considered as the standard approach to describe prokaryotic species, polyphasic taxonomy frequently employs information other than 16S rRNA gene sequences to describe novel species (Brenner et al., 2005). We contend that this approach is reasonable and reckon that the same approach is applicable to the classification of multicellular eukaryotes as well. In our opinion, a phylogenetic backbone, instead of phenetic clustering or otherwise, provides an objective framework of a taxonomy by illuminating the plethora of phenotypic characteristic with evolutionary insight along the nodes and branches of the phylogenetic tree; the phylogenetic tree, however, does not directly translates itself into a taxonomic nomenclature precisely because of the varying rates of divergence and evolution. The value of a polyphasic taxonomy, presumably one that is firmly grounded on a properly inferred phylogeny, is the necessary integration of subjective information to derive cut-offs that make a phylogenetic tree meaningful not to its optimization algorithm but to the scientists' minds. To this end, cladistics provide the stable, objective framework which is negligent (Mayr, 1974) and therefore immune (Page and Holmes, 1998) to the varying tempi of evolution.

CHANGING TAXONOMIES: CONFLICT AMONG THE WARLORDS

The three-domain classification has not been unanimously agreed since its proposal. With the advancement of cytology and the extensive use of molecular phylogenetics, taxonomists were obliged to change their opinion on kingdom delimitation from time to time. Parallel to Woese's work, Cavalier-Smith proposed a system of nine kingdoms (Cavalier-Smith, 1981), subsequently one with seven kingdoms (Cavalier-Smith, 1986), then another with eight kingdoms and 10 subkingdoms (Cavalier-Smith, 1993) and a more recent one with only six kingdoms (Cavalier-Smith, 1998, 2004). Notably, the Cavalier-Smith school did not subscribe to the splitting of Kingdom *Prokaryote* and the new Domain concept and instead proposed the use of below-kingdom hierarchies such as *Subkingdom*, *Infrakingdom* and *Parvkingdom*. Stephen Jay Gould (1943–2002), in accord with Cavalier-Smith, commented that the renaming, reclassification and elevation of *Archaea* as a taxonomic rank above that of a Kingdom "grossly inflates the importance of the differences between the two kingdoms *Eubacteria* and *Archaebacteria*" (Cavalier-Smith, 1993; Margulis, et al., 2009).

We believe that whether all prokaryotic organisms should be classified into a single kingdom is merely subjective and conceptual in the absence of a credible phylogeny. If the phylogenetic tree and sequence of divergence depicted in Woese's initial proposal are true, i.e. the *Bacteria-Archaea* split occurred before the *Archaea-Eukaryota* split (Iwabe et al., 1989; Cammarano et al., 1992; Brown and Doolittle, 1995; Yang and Roberts, 1995; Baldauf et al., 1996; Gribaldo and Cammarano, 1998; Woese, 2000; Lake et al., 2009; Fournier and Gogarten, 2010) instead of erroneous (Saccone et al., 1995; Lawson et al., 1996; Forterre and Philippe, 1999; Lopez et al., 1999; Philippe and Forterre, 1999; Caetano-Anolles, 2002; Gribaldo and Philippe, 2002; Bapteste and Brochier, 2004; Lake et al., 2008; Sun and Caetano-Anolles, 2009; Dagan et al., 2010;

Protein & Cell

Valas and Bourne, 2011) or indeterminate due to horizontal gene transfer (Lake and Rivera, 2004; Bapteste and Walsh, 2005), then to unite the prokaryotes into a hierarchy excluding *Eukaryota* would create paraphyly. On the other hand, to apply such cladistic concepts to a tree of life is to assume that a strictly bifurcating phylogeny existed for the early evolutionary history despite the potentially extensive horizontal gene transfers (Doolittle and Brown, 1994). Apart from alternative roots of the universal tree of life, proponents of a single prokaryotic kingdom (or superkingdom) have also argued by current molecular evidence and highlighted the many reproductive, organizational, developmental and nutritional differences between prokaryotes and eukaryotes—which, according to the proponents, represent whole-organism criteria instead of partial phylogenies. Such polarized debates have even resulted in suggestions of abandoning the use of some common terms. The "bacterial chromosome", for example, is "used by molecular biologists to refer to the *genophore*, is confusing and its use should be avoided" (Margulis, et al., 2009). We note, however, the terminology suggested by Margulis and Chapmanto mainly served to highlight the differences between the *Archaea* and *Eukaryota*; and their repeated references to the methanogenic and halophilic archaea as being "bacteria" is perhaps more confusing to many microbiologists.

On the other front, while the division of the prokaryotes into new domains or supra-kingdom hierarchies has not yet been completely settled, cladistics has penetrated the subdomain hierarchies in *Eukaryota*. One of the more illuminating examples occurred at the interface of the eukaryotic clades, the protists. Although still considered as an important grade of classification by certain pheneticists (Cavalier-Smith, 2004), the convenience of classification once thought to be an indispensible virtue of this paraphyletic taxon has gone into disfavor as monophyletic groups became the phylogenetic ideal (Simpson and Roger, 2004). With the advent of phylogenetics, it was recognized that the classification of eukaryotic organisms—especially those that could only be classified based on mainly phenetic characters—should follow that of prokaryotic taxonomy to include gene trees and multi-gene phylogenies for enhanced validity (Keeling et al., 2005). The extensive revisions brought to protist taxonomy by molecular phylogenetics aggravated the incompatibilities between the International Code of Zoological Nomenclature and International Code of Botanical Nomenclature to the point that neither was considered adequate for a practical taxonomy (Adl et al., 2007). The paraphyletic Kingdom *Protoctista*, although still included in many junior and college science curricula, has also been noted as being a "phylogenetic dustbin" (Bradfield, 2009). In through and through cladistic-based databases such as the NCBI Taxonomy database, it is perhaps reasonable to see *Protoctista* not being recognized as a valid taxonomic rank (Federhen, 2012).

THE PHYLOGENOMIC ERA: END OF THE THREE DOMAINS?

Phylogenetic analysis in its dawn was limited to one or a few

selected loci, which precluded rigorous methods to weight and evaluate conflicting topologies even if present. Very often, the additional data gained from sequencing alternate genetic loci were non-selectively integrated by concatenation into an alignment of an established locus in the hope of giving better resolution or reliability to the resulting phylogenetic tree. Indeed, when software-based methodologies for alignment trimming were first presented (Castresana, 2000), the removal of highly variable and unreliable sequence positions was met with reluctance since the reduced information content of the trimmed alignment occasionally concerned the researcher (Talavera and Castresana, 2007). In the age of high-throughput DNA sequencing, ironically, one of the major issues emerged as the selection of the more reliable loci among the plethora of targets in the sequenced genome (Salichos and Rokas, 2013).

With whole-genome comparative genomics (Razin, 1997) and the rise of phylogenomics (Sicheritz-Ponten and Andersson, 2001), the dynamic nature of genomes is being increasingly recognized especially in the most sequenced domains, the *Bacteria* and *Archaea*. The long-standing debate of whether horizontal gene transfer has compromised the universal tree of life once again became relevant, not because of the then paucity of targets but now due to the abundance of putative horizontal gene transfers identified. While early studies have supported the separation of the prokaryotic domains (Sicheritz-Ponten and Andersson, 2001), more recent complete genome comparisons have revealed extensive evidence for horizontal genetic transfers in the history of bacterial and archaeal evolution (Koonin and Wolf, 2008). Coupling to the pervasive gene loss and functional plasticity of many genes, consistent phylogenetic signals can only be observed at relatively shallow phylogenetic depths and the concept of a phylogenetic tree may not be strictly applicable especially at the base of the tree where the domains diverge (Puigbo et al., 2009).

We envisage that there may soon be another debate which ultimately challenges the central tenets of cladistics. Supporters of the tree of life concept have devised novel methods to statistically reconcile the differing topologies of phylogenetic trees from different genes and recover a species tree (Abby et al., 2010; Schliep et al., 2011). Although it has also been suggested that the more efficient way to estimate the common phylogeny, i.e. the species tree, underlying the different genes sampled is the concatenation approach given proper consideration for among-gene heterogeneity (Yang and Rannala, 2012), this proposal has also been questioned based on the total incongruence among all genes observed in a large-scale phylogenomic study (Salichos and Rokas, 2013). Regardless of the actual results obtained, these methodological inconsistencies may render the tree of life proposal more prone to attack by the web of life supporters (Kunin et al., 2005), especially when the methods of phylogenetic and phylogenomic network reconstruction become mature and widely available (Woolley et al., 2008).

Why should this challenge the clade concept? If one considers the basis of Hennigian cladistics, it is clear that its

foundation is based on the concept of homology (Hennig, 1950, 1966). Without proper ascertainment of homology, the comparison of apomorphic characters cannot be performed. Clearly, homology since the early days of molecular phylogenetics has been considered an all-or-none phenomenon with no intermediate or fractional levels (Theißen, 2002). The extensive and on-going horizontal gene transfers therefore centrally challenge the concept of homology at the organism level far beyond the within-species sexual reproduction or the occasional hybridization between species. Even if we neglect viral evolution, which is exactly characterized by an abundance of recombination (Woo et al., 2009; Lau et al., 2010, 2011b, 2011c; Yip et al., 2011), for they may not be entirely representative of the membrane-bound evolutionary pathways taken by the cellular organisms, the mosaic nature of the eukaryotic (Golding and Gupta, 1995; Gupta, 1998; Ribeiro and Golding, 1998) and even archaeal genomes (Kennedy et al., 2001; Deppenmeier et al., 2002) is disturbing to the cladists. Given that recombination occurs between "homologous" regions in a genome and foreign genetic elements can invade and integrate into existing genes, the absolute demarcation of homology may be compromised even at the gene level. Do clades stand, absolutely, as a phylogenetic backbone supporting an essentially tree-like phylogeny to the point that we can regard the horizontal gene transfer events as signal rather than noise (Abby et al., 2012)? Alternatively, should we modify our understanding of homology and clades, updating our traditional methodologies (McDade, 1990, 1992) to include a probabilistic element to recognize the non-major, alternative contributions to the evolutionary history of a gene and an organism (Posada and Crandall, 2001)? Or, have the ideas of homology and clades been rendered obsolete and novel concepts should be introduced to help us understand the somehow reticulated evolutionary history of the domains of life?

CONCLUDING REMARKS

In this review, we examined the transforming influences of cladistics on the classification and taxonomy of cellular organisms. These influences were not the sole result of the cladistic concepts. The advent of molecular biology techniques led to the widespread availability of genetic data, which allowed the geneology of organisms to be inferred using the concurrently developed methodologies, algorithms and, perhaps most importantly, modern and affordable computers. The development of high-throughput sequencing in the last decade has led to a quantum leap in the availability of genetic data as the complete repertoire of genes of microbes and even animals and plants became all at once available.

It is a pity that neither Charles Darwin (1809–1882) nor Willi Hennig (1913–1976) lived to witness these changes. With the power of the high-throughput sequencing technologies, for the first time in history, we are beginning to see the overwhelming diversity of genomic architecture, magnitude of gene flow and genetic variability among extent or even extinct organisms.

Then again, less optimistically, one may doubt if current taxon sampling, methodologies and computational power allow the phylogenetic ideals of the cladists—or perhaps the practicality-driven taxonomy of the pheneticists—to be found, since even with genomic understanding the functional adaptations of organisms to their specific niches are not immediately apparent.

We opt to be optimistic. With the recent success in modeling an entire organism by using its complete genetic information (Karr et al., 2012), we believe that the gap between evolutionary simulations at the genetic and genomic level (Edgar et al., 2013) and the phenotype or adaptability level will eventually be bridged. The debates that we have seen and predicted, however, will not be settled by such as taxonomic classifications serve the human scientist. With sympathy, we quote the respected *Bergey's Manual* in our field to end this review, with the hope that it may serve as a reminder to the wider taxonomic audience:

"…bacterial classifications are devised for microbiologists, not for the entities being classified. Bacteria show little interest in the matter of their classification. For the taxonomist, this is sometimes a very sobering thought!" (Brenner et al., 2005).

COMPLIANCE WITH ETHICS GUIDELINES

Chi-Chun Ho, Susanna K P Lau and Patrick C Y Woo declare no conflict of interest. This review article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

(1611). THE HOLY BIBLE, Conteyning the Old Testament, AND THE NEW: Newly Translated out of the Originall tongues: & with the former Translations diligently compared and revised, by his Majesties special Comandement. Appointed to be read in Churches. (London, Robert Barker).

Abby, S.S., Tannier, E., Gouy, M., and Daubin, V. (2010). Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests. BMC Bioinformatics 11, 324.

Abby, S.S., Tannier, E., Gouy, M., and Daubin, V. (2012). Lateral gene transfer as a support for the tree of life. Proc Natl Acad Sci U S A 109, 4962–4967.

Achenbach-Richter, L., Gupta, R., Zillig, W., and Woese, C.R. (1988). Rooting the archaebacterial tree: the pivotal role of *Thermococcus celer* in archaebacterial evolution. Syst Appl Microbiol 10, 231–240.

Adl, S.M., Leander, B.S., Simpson, A.G., Archibald, J.M., Anderson, O.R., Bass, D., Bowser, S.S., Brugerolle, G., Farmer, M.A., Karpov, S., et al. (2007). Diversity, nomenclature, and taxonomy of protists. Syst Biol 56, 684–689.

Ashlock, P.D. (1974). The uses of cladistics. Annu Rev Ecol Systemat 5, 81–99.

Protein & Cell

Protein & Cell

Avise, J.C. (1974). Systematic value of electrophoretic data. Syst Biol 23, 465–481.

Balch, W.E., Magrum, L.J., Fox, G.E., Wolfe, R.S., and Woese, C.R. (1977). An ancient divergence among the bacteria. J Mol Evol 9, 305–311.

Baldauf, S.L., Palmer, J.D., and Doolittle, W.F. (1996). The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc Natl Acad Sci U S A 93, 7749–7754.

Bapteste, E., and Brochier, C. (2004). On the conceptual difficulties in rooting the tree of life. Trends Microbiol 12, 9–13.

Bapteste, E., and Walsh, D.A. (2005). Does the 'Ring of Life' ring true? Trends Microbiol 13, 256–261.

Boyden, A. (1947). Homology and analogy. a critical review of the meanings and implications of these concepts in biology. Amer Mid Natur 37, 648–669.

Branfield, P. and Potter, S. (2009). Edexcel IGCSE biology (Harlow, Pearson).

Brenner, D., Staley, J., and Krieg, N. (2005). Classification of procaryotic organisms and the concept of bacterial speciation. In Bergey's manual of systematic bacteriology, Brenner, D., Krieg, N., Staley, J., and Garrity, G. eds. (Springer US), pp. 27–32.

Bridge, P.D., and Sneath, P.H. (1983). Numerical taxonomy of *Streptococcus*. J Gen Microbiol 129, 565–597.

Broom, A., and Sneath, P.H. (1981). Numerical taxonomy of *Haemophilus*. J Gen Microbiol 126, 123–149.

Brown, J.R., and Doolittle, W.F. (1995). Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci U S A 92, 2441–2445.

Brundin, L. (1966). Transantarctic relationships and their significance, as evidenced by chironomid midges. With a monograph of the subfamilies *Podonominae* and *Aphroteniinae* and the austral *Heptagyiae* (Stockholm, Almqvist & Wiksell).

Burggraf, S., Fricke, H., Neuner, A., Kristjansson, J., Rouvier, P., Mandelco, L., Woese, C.R., and Stetter, K.O. (1990). *Methanococcus igneus* sp. nov., a novel hyperthermophilic methanogen from a shallow submarine hydrothermal system. Syst Appl Microbiol 13, 263–269.

Burkholder, J.M., and Glasgow, H.B., Jr. (1997). Trophic controls on stage transformations of a toxic ambush-predator dinoflagellate. J Eukaryot Microbiol 44, 200–205.

Caetano-Anolles, G. (2002). Evolved RNA secondary structure and the rooting of the universal tree of life. J Mol Evol 54, 333–345.

Cammarano, P., Palm, P., Creti, R., Ceccarelli, E., Sanangelantoni, A.M., and Tiboni, O. (1992). Early evolutionary relationships among known life forms inferred from elongation factor EF-2/EF-G sequences: phylogenetic coherence and structure of the archaeal domain. J Mol Evol 34, 396–405.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17, 540–552.

Cavalier-Smith, T. (2004). Only six kingdoms of life. Proc Biol Sci 271, 1251–1262.

Chan, J.F., Lau, S.K., Curreem, S.O., To, K.K., Leung, S.S., Cheng, V.C., Yuen, K.Y., and Woo, P.C. (2012). First report of spontaneous intrapartum *Atopobium vaginae* bacteremia. J Clin Microbiol 50, 2525–2528.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J.,

Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37, D141–145.

Colwell, R.R. (1970). Polyphasic taxonomy of the genus vibrio: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species. J Bacteriol 104, 410–433.

Dagan, T., Roettger, M., Bryant, D., and Martin, W. (2010). Genome networks root the tree of life between prokaryotic domains. Genome Biol Evol 2, 379–392.

Darwin, C. (1859). On the origin of species by means of natural selection (London,, J. Murray).

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., et al. (2002). The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J Mol Microbiol Biotechnol 4, 453–461.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72, 5069–5072.

Doolittle, W.F., and Brown, J.R. (1994). Tempo, mode, the progenote, and the universal root. Proc Natl Acad Sci U S A 91, 6721–6728.

Edgar, R., Asimenos, G., Batzoglou, S., and Sidow, A. (2013). Evolver: a whole-genome sequence evolution simulator.

Federhen, S. (2012). The NCBI Taxonomy database. Nucleic Acids Res 40, D136–143.

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R., and Raff, R.A. (1988). Molecular phylogeny of the animal kingdom. Science 239, 748–753.

Forterre, P., and Philippe, H. (1999). Where is the root of the universal tree of life? Bioessays 21, 871–879.

Fournier, G.P., and Gogarten, J.P. (2010). Rooting the ribosomal tree of life. Mol Biol Evol 27, 1792–1801.

Fox, G.E., Magrum, L.J., Balch, W.E., Wolfe, R.S., and Woese, C.R. (1977). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. Proc Natl Acad Sci U S A 74, 4537–4541.

Golding, G.B., and Gupta, R.S. (1995). Protein-based phylogenies support a chimeric origin for the eukaryotic genome. Mol Biol Evol 12, 1–6.

Graur, D., and Li, W.-H. (2000). Fundamentals of molecular evolution, 2nd edn (Sunderland, Massachusetts, USA., Sinauer Associates).

Gribaldo, S., and Cammarano, P. (1998). The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. J Mol Evol 47, 508–516.

Gribaldo, S., and Philippe, H. (2002). Ancient phylogenetic relationships. Theor Popul Biol 61, 391–408.

Gupta, R., Lanter, J.M., and Woese, C.R. (1983). Sequence of the 16S ribosomal RNA from *Halobacterium volcanii*, an archaebacterium. Science 221, 656–659.

Gupta, R.S. (1998). Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol Mol Biol Rev 62, 1435–1491.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., Richardson, P.M., and DeLong, E.F. (2006). Genomic analysis of the uncultivated marine

crenarchaeote *Cenarchaeum symbiosum*. Proc Natl Acad Sci U S A 103, 18296–18301.

Hennig, W. (1950). Grundzüge einer Theorie der phylogenetischen Systematik (Berlin, Deutscher zentralverlag).

Hennig, W. (1965). Phylogenetic Systematics. Annu Rev Entomol 10, 97–116.

Hennig, W. (1966). Phylogenetic systematics (Urbana,, University of Illinois Press).

Ho, C.C., Wu, A.K., Tse, C.W., Yuen, K.Y., Lau, S.K., and Woo, P.C. (2012). Automated pangenomic analysis in target selection for PCR detection and identification of bacteria by use of ssGeneFinder Webserver and its application to *Salmonella enterica* serovar Typhi. J Clin Microbiol 50, 1905–1911.

Ho, C.C., Yuen, K.Y., Lau, S.K., and Woo, P.C. (2011). Rapid identification and validation of specific molecular targets for detection of *Escherichia coli* O104:H4 outbreak strain by use of high-throughput sequencing data from nine genomes. J Clin Microbiol 49, 3714–3716.

Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Taylor, T.D., Kudo, T., Sakaki, Y., Toyoda, A., Hattori, M., and Ohkuma, M. (2008a). Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. Proc Natl Acad Sci U S A 105, 5555–5560.

Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Toh, H., Taylor, T.D., Kudo, T., Sakaki, Y., Toyoda, A., Hattori, M., et al. (2008b). Genome of an endosymbiont coupling N2 fixation to cellulolysis within protist cells in termite gut. Science 322, 1108–1109.

Hubbs, C.L. (1944). Concepts of homology and analogy. Amer Nat 78, 289–307.

Huxley, J. (1957). The three types of evolutionary process. Nature 180, 454–455.

Huxley, J. (1959). Clades and grades. in function and taxonomic importance: a symposium. Cain, A.J. ed. (London, Systematics Association).

Hyer, B.H., McCarthy, B.J., and Bolton, E.T. (1964). A molecular approach in the systematics of higher organisms. dna interactions provide a basis for detecting common polynucleotide sequences among diverse organisms. Science 144, 959–967.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci U S A 86, 9355–9359.

Johnson, M.L., and Wicks, M.J. (1959). Serum Protein Electrophoresis in Mammals—Taxonomic Implications. Syst Biol 8, 88–95.

Jones, J.H., Card, W., Chapman, M., Lennard-Jones, J.E., Morson, B.C., Sackin, M.J., and Sneath, P.H. (1970). The application of numerical taxonomy to the separation of cllonic inflammatory disease. Gut 11, 1062.

Jones, J.H., Lennard-Jones, J.E., Morson, B.C., Chapman, M., Sackin, M.J., Sneath, P.H., Spicer, C.C., and Card, W.I. (1973). Numerical taxonomy and discriminant analysis applied to non-specific colitis. Q J Med 42, 715–732.

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from genotype. Cell 150, 389–401.

Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., Pearlman, R.E., Roger, A.J., and Gray, M.W. (2005). The tree of eukaryotes. Trends Ecol Evol 20, 670–676.

Kennedy, S.P., Ng, W.V., Salzberg, S.L., Hood, L., and DasSarma, S. (2001). Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. Genome Res 11, 1641–1650.

Koonin, E.V., and Wolf, Y.I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res 36, 6688–6719.

Kunin, V., Goldovsky, L., Darzentas, N., and Ouzounis, C.A. (2005). The net of life: reconstructing the microbial phylogenetic network. Genome Res 15, 954–959.

Lake, J.A., and Rivera, M.C. (2004). Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction. Mol Biol Evol 21, 681–690.

Lake, J.A., Servin, J.A., Herbold, C.W., and Skophammer, R.G. (2008). Evidence for a new root of the tree of life. Syst Biol 57, 835–843.

Lake, J.A., Skophammer, R.G., Herbold, C.W., and Servin, J.A. (2009). Genome beginnings: rooting the tree of life. Philos. Trans R Soc Lond B Biol Sci 364, 2177–2185.

Lau, S.K., Curreem, S.O., Ngan, A.H., Yeung, C.K., Yuen, K.Y., and Woo, P.C. (2011a). First report of disseminated *Mycobacterium* skin infections in two liver transplant recipients and rapid diagnosis by hsp65 gene sequencing. J Clin Microbiol 49, 3733–3738.

Lau, S.K., Lee, P., Tsang, A.K., Yip, C.C., Tse, H., Lee, R.A., So, L.Y., Lau, Y.L., Chan, K.H., Woo, P.C., et al. (2011b). Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. J Virol 85, 11325–11337.

Lau, S.K., Li, K.S., Huang, Y., Shek, C.T., Tse, H., Wang, M., Choi, G.K., Xu, H., Lam, C.S., Guo, R., et al. (2010). Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. J Virol 84, 2808–2819.

Lau, S.K., Ng, K.H., Woo, P.C., Yip, K.T., Fung, A.M., Woo, G.K., Chan, K.M., Que, T.L., and Yuen, K.Y. (2006a). Usefulness of the Micro-Seq 500 16S rDNA bacterial identification system for identification of anaerobic Gram positive bacilli isolated from blood cultures. J Clin Pathol 59, 219–222.

Lau, S.K., Woo, P.C., Chan, C.Y., Woo, W.L., Woo, G.K., and Yuen, K.Y. (2005). Typhoid fever associated with acute appendicitis caused by an H1-j strain of *Salmonella enterica* serotype Typhi. J Clin Microbiol 43, 1470–1472.

Lau, S.K., Woo, P.C., Luk, W.K., Fung, A.M., Hui, W.T., Fong, A.H., Chow, C.W., Wong, S.S., and Yuen, K.Y. (2006b). Clinical isolates of *Streptococcus iniae* from Asia are more mucoid and beta-hemolytic than those from North America. Diagn Microbiol Infect Dis 54, 177–181.

Lau, S.K., Woo, P.C., Yip, C.C., Li, K.S., Fu, C.T., Huang, Y., Chan, K.H., and Yuen, K.Y. (2011c). Co-existence of multiple strains of two novel porcine bocaviruses in the same pig, a previously undescribed phenomenon in members of the family Parvoviridae, and evidence for inter- and intra-host genetic diversity and recombination. J Gen Virol 92, 2047–2059.

Lawson, F.S., Charlebois, R.L., and Dillon, J.A. (1996). Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which

Protein & Cell

can root the tree of life. Mol Biol Evol 13, 970–977.

Li, W.H., and Tanimura, M. (1987). The molecular clock runs more slowly in man than in apes and monkeys. Nature 326, 93–96.

Linnaeus, C. (1751). Philosophia botanica.

Lopez, P., Forterre, P., and Philippe, H. (1999). The root of the tree of life in the light of the covarion model. J Mol Evol 49, 496–508.

Ludwig, W., and Klenk, H.-P. (2005). Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics. In Bergey's manual of systematic bacteriology, Brenner, D., Krieg, N., Staley, J., and Garrity, G. eds. (Springer US), pp. 49–66.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., et al. (2004). ARB: a software environment for sequence data. Nucleic Acids Res 32, 1363–1371.

Magrum, L.J., Luehrsen, K.R., and Woese, C.R. (1978). Are extreme halophiles actually "bacteria"? J Mol Evol 11, 1–8.

Margoliash, E., Smith, E.L., Kreil, G., and Tuppy, H. (1961). Amino-acid sequence of horse heart cytochrome c. Nature 192, 1125–1127.

Mayr, E. (1974). Cladistic analysis or cladistic classification. Zeitschrift für zoologische Systematik und Evolutionsforschung 12, 94–128.

Mayr, E. (1981). Biological classification: toward a synthesis of opposing methodologies. Science 214, 510–516.

McCarthy, B.J., and Bolton, E.T. (1963). An approach to the measurement of genetic relatedness among organisms. Proc Natl Acad Sci U S A 50, 156–164.

McDade, L. (1990). Hybrids and phylogenetic systematics i. patterns of character expression in hybrids and their implications for cladistic analysis. Evolution 44, 1685–1700.

McDade, L.A. (1992). Hybrids and phylogenetic systematics ii. the impact of hybrids on cladistic analysis. Evolution 46, 1329–1346.

Medlin, L., Elwood, H.J., Stickel, S., and Sogin, M.L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71, 491–499.

Myers, G.S. (1952). The nature of systematic biology and of a species description. Syst Zool 1, 106–111.

Noller, H.F., and Woese, C.R. (1981). Secondary structure of 16S ribosomal RNA. Science 212, 403–411.

Olsen, G.J., Overbeek, R., Larsen, N., Marsh, T.L., McCaughey, M.J., Maciukenas, M.A., Kuan, W.M., Macke, T.J., Xing, Y., and Woese, C.R. (1992). The ribosomal database project. Nucleic Acids Res 20 Suppl, 2199–2200.

Olsen, G.J., Pace, N.R., Nuell, M., Kaine, B.P., Gupta, R., and Woese, C.R. (1985). Sequence of the 16S rRNA gene from the thermoacidophilic archaebacterium *Sulfolobus solfataricus* and its evolutionary implications. J Mol Evol 22, 301–307.

Oyaizu, H., Debrunner-Vossbrinck, B., Mandelco, L., Studier, J.A., and Woese, C.R. (1987). The green non-sulfur bacteria: A deep branching in the eubacterial line of descent. Syst Appl Microbiol 9, 47–53.

Oyaizu, H., and Woese, C.R. (1985). Phylogenetic relationships among the sulfate respiring bacteria, myxobacteria and purple bacteria. Syst Appl Microbiol 6, 257–263.

Page, R.D.M., and Holmes, E.C. (1998). Molecular evolution: a phylogenetic approach (Oxford ; Malden, MA, Blackwell Science).

Paster, B.J., Ludwig, W., Weisburg, W.G., Stackebrandt, E., Hespell, R.B., Hahn, C.M., Reichenbach, H., Stetter, K.O., and Woese, C.R. (1985). A phylogenetic grouping of the *Bacteroides*, *Cytophagas*, and certain *Flavobacteria*. Syst Appl Microbiol 6, 34–42.

Paster, B.J., Stackebrandt, E., Hespell, R.B., Hahn, C.M., and Woese, C.R. (1984). The phylogeny of the spirochetes. Syst Appl Microbiol 5, 337–351.

Philippe, H., and Forterre, P. (1999). The rooting of the universal tree of life is not reliable. J Mol Evol 49, 509–523.

Posada, D., and Crandall, K.A. (2001). Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol 16, 37–45.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35, 7188–7196.

Puigbo, P., Wolf, Y.I., and Koonin, E.V. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. J Biol 8, 59.

Razin, S. (1997). Comparative genomics of mycoplasmas. Wien Klin Wochenschr 109, 551–556.

Rensch, B. (1954). Neuere Probleme der Abstammungslehre: die transspezifische Evolution, 2., stark veränderte Aufl. edn (Stuttgart, F. Enke).

Ribeiro, S., and Golding, G.B. (1998). The mosaic nature of the eukaryotic nucleus. Mol Biol Evol 15, 779–788.

Saccone, C., Gissi, C., Lanave, C., and Pesole, G. (1995). Molecular classification of living organisms. J Mol Evol 40, 273–279.

Salichos, L., and Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331.

Sanger, F., Donelson, J.E., Coulson, A.R., Kossel, H., and Fischer, D. (1973). Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage fl DNA. Proc Natl Acad Sci U S A 70, 1209–1213.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74, 5463–5467.

Schliep, K., Lopez, P., Lapointe, F.J., and Bapteste, E. (2011). Harvesting evolutionary signals in a forest of prokaryotic gene trees. Mol Biol Evol 28, 1393–1405.

Sicheritz-Ponten, T., and Andersson, S.G. (2001). A phylogenomic approach to microbial evolution. Nucleic Acids Res 29, 545–552.

Simpson, A.G.B., and Roger, A.J. (2004). The real 'kingdoms' of eukaryotes. Curr Biol 14, R693–R696.

Sneath, P.H., and Sokal, R.R. (1962). Numerical taxonomy. Nature 193, 855–860.

Sneath, P.H., Stevens, M., and Sackin, M.J. (1981). Numerical taxonomy of *Pseudomonas* based on published records of substrate utilization. Antonie van Leeuwenhoek 47, 423–448.

Sneath, P.H.A. (1995). Thirty years of numerical taxonomy. Syst Biol 44, 281–298.

Sokal, R.R.S.P.H.A. (1963). Principles of numerical taxonomy (San Francisco, Freeman.).

Sun, F.J., and Caetano-Anolles, G. (2009). The evolutionary history of the structure of 5S ribosomal RNA. J Mol Evol 69, 430–443.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56, 564–577.

Theißen, G. (2002). Orthology: secret life of genes. Nature 415, 741–741.

Titsworth, E., Grunberg, E., Beskid, G., Cleeland, R., Jr., and Delorenzo, W.F. (1969). Efficiency of a multitest system (Enterotube) for rapid identification of Enterobacteriaceae. Appl Microbiol 18,

Protein & Cell

207–213.

Tse, H., Tsang, A.K., Tsoi, H.W., Leung, A.S., Ho, C.C., Lau, S.K., Woo, P.C., and Yuen, K.Y. (2012). Identification of a novel bat papillomavirus by metagenomics. PLoS ONE 7, e43986.

US National Library of Medicine (2006). Genetic Speciation.

Valas, R.E., and Bourne, P.E. (2011). The origin of a derived superkingdom: how a gram-positive bacterium crossed the desert to become an archaeon. Biol Direct 6, 16.

Washington, J.A., 2nd, Yu, P.K., and Martin, W.J. (1971). Evaluation of accuracy of multitest micromethod system for identification of *Enterobacteriaceae*. Appl Microbiol 22, 267–269.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., and Lane, D.J. (1991). 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173, 697–703.

Weisburg, W.G., Giovannoni, S.J., and Woese, C.R. (1989). The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. Syst Appl Microbiol 11, 128–134.

Williams, D.M., and Ebach, M.C. (2009). What, exactly, is cladistics? Re-writing the history of systematics and biogeography. Acta Biotheor 57, 249–268.

Woese, C.R. (1979). A proposal concerning the origin of life on the planet earth. J Mol Evol 13, 95–101.

Woese, C.R. (2000). Interpreting the universal phylogenetic tree. Proc Natl Acad Sci U S A 97, 8392–8396.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74, 5088–5090.

Woese, C.R., Gupta, R., Hahn, C.M., Zillig, W., and Tu, J. (1984a). The phylogenetic relationships of three sulfur dependent archaebacteria. Syst Appl Microbiol 5, 97–105.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990a). Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. Proc Natl Acad Sci U S A 87, 4576–4579.

Woese, C.R., Magrum, L.J., and Fox, G.E. (1978). *Archaebacteria*. J Mol Evol 11, 245–251.

Woese, C.R., Magrum, L.J., Gupta, R., Siegel, R.B., Stahl, D.A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J.J., et al. (1980a). Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. Nucleic Acids Res 8, 2275–2293.

Woese, C.R., Maloy, S., Mandelco, L., and Raj, H.D. (1990b). Phylogenetic placement of the *Spirosomaceae*. Syst Appl Microbiol 13, 19–23.

Woese, C.R., Mandelco, L., Yang, D., Gherna, R., and Madigan, M.T. (1990c). The case for relationship of the *flavobacteria* and their relatives to the green sulfur bacteria. Syst Appl Microbiol 13, 258–262.

Woese, C.R., Maniloff, J., and Zablen, L.B. (1980b). Phylogenetic analysis of the mycoplasmas. Proc Natl Acad Sci U S A 77, 494–498.

Woese, C.R., Stackebrandt, E., Weisburg, W.G., Paster, B.J., Madigan, M.T., Fowler, V.J., Hahn, C.M., Blanz, P., Gupta, R., Nealson, K.H., et al. (1984b). The phylogeny of purple bacteria: The alpha subdivision. Syst Appl Microbiol 5, 315–326.

Woese, C.R., Weisburg, W.G., Hahn, C.M., Paster, B.J., Zablen, L.B., Lewis, B.J., Macke, T.J., Ludwig, W., and Stackebrandt, E. (1985). The phylogeny of purple bacteria: The gamma subdivision. Syst Appl Microbiol 6, 25–33.

Woese, C.R., Yang, D., Mandelco, L., and Stetter, K.O. (1990d). The

flexibacter-flavobacter connection. Syst Appl Microbiol 13, 161–165.

Woo, P.C., Chong, K.T., Tse, H., Cai, J.J., Lau, C.C., Zhou, A.C., Lau, S.K., and Yuen, K.Y. (2006). Genomic and experimental evidence for a potential sexual cycle in the pathogenic thermal dimorphic fungus *Penicillium marneffei*. FEBS Lett 580, 3409–3416.

Woo, P.C., Fung, A.M., Wong, S.S., Tsoi, H.W., and Yuen, K.Y. (2001a). Isolation and characterization of a *Salmonella enterica* serotype Typhi variant and its clinical and public health implications. J Clin Microbiol 39, 1190–1194.

Woo, P.C., Lam, C.W., Tam, E.W., Leung, C.K., Wong, S.S., Lau, S.K., and Yuen, K.Y. (2012). First discovery of two polyketide synthase genes for mitorubrinic acid and mitorubrinol yellow pigment biosynthesis and implications in virulence of *Penicillium marneffei*. PLoS Negl Trop Dis 6, e1871.

Woo, P.C., Lau, S.K., Huang, Y., and Yuen, K.Y. (2009). Coronavirus diversity, phylogeny and interspecies jumping. Exp Biol Med (Maywood) 234, 1117–1127.

Woo, P.C., Lau, S.K., Teng, J.L., Tse, H., and Yuen, K.Y. (2008). Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect 14, 908–934.

Woo, P.C., Lau, S.K., Woo, G.K., Fung, A.M., Ngan, A.H., Hui, W.T., and Yuen, K.Y. (2003a). Seronegative bacteremic melioidosis caused by *Burkholderia pseudomallei* with ambiguous biochemical profile: clinical importance of accurate identification by 16S rRNA gene and *groEL* gene sequencing. J Clin Microbiol 41, 3973–3977.

Woo, P.C., Leung, P.K., Leung, K.W., and Yuen, K.Y. (2000). Identification by 16S ribosomal RNA gene sequencing of an *Enterobacteriaceae* species from a bone marrow transplant recipient. Mol Pathol 53, 211–215.

Woo, P.C., Leung, P.K., Wong, S.S., Ho, P.L., and Yuen, K.Y. (2001b). *groEL* encodes a highly antigenic protein in *Burkholderia pseudomallei*. Clin Diagn Lab Immunol 8, 832–836.

Woo, P.C., Ng, K.H., Lau, S.K., Yip, K.T., Fung, A.M., Leung, K.W., Tam, D.M., Que, T.L., and Yuen, K.Y. (2003b). Usefulness of the MicroSeq 500 16S ribosomal DNA-based bacterial identification system for identification of clinically significant bacterial isolates with ambiguous biochemical profiles. J Clin Microbiol 41, 1996–2001.

Woo, P.C., Tam, E.W., Chong, K.T., Cai, J.J., Tung, E.T., Ngan, A.H., Lau, S.K., and Yuen, K.Y. (2010). High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in *Penicillium marneffei*. FEBS J. 277, 3750–3758.

Woo, P.C., Teng, J.L., Yeung, J.M., Tse, H., Lau, S.K., and Yuen, K.Y. (2011). Automated identification of medically important bacteria by 16S rRNA gene sequencing using a novel comprehensive database, 16SpathDB. J Clin Microbiol 49, 1799–1809.

Woo, P.C., Wong, S.S., Lum, P.N., Hui, W.T., and Yuen, K.Y. (2001c). Cell-wall-deficient bacteria and culture-negative febrile episodes in bone-marrow-transplant recipients. Lancet 357, 675–679.

Woo, P.C., Zhen, H., Cai, J.J., Yu, J., Lau, S.K., Wang, J., Teng, J.L., Wong, S.S., Tse, R.H., Chen, R., et al. (2003c). The mitochondrial genome of the thermal dimorphic fungus *Penicillium marneffei* is more closely related to those of molds than yeasts. FEBS Lett 555, 469–477.

Woolley, S.M., Posada, D., and Crandall, K.A. (2008). A comparison of phylogenetic network methods using computer simulation. PLoS

Protein & Cell

ONE 3, e1913.

Yang, D., Kaine, B.P., and Woese, C.R. (1985a). The phylogeny of *Archaebacteria*. Syst Appl Microbiol 6, 251–256.

Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J., and Woese, C.R. (1985b). Mitochondrial origins. Proc Natl Acad Sci U S A 82, 4443–4447.

Yang, D., and Woese, C.R. (1989). Phylogenetic sructure of the "*Leuconostocs*": an interesting case of a rapidly evolving organism. Syst Appl Microbiol 12, 145–149.

Yang, Z., and Rannala, B. (2012). Molecular phylogenetics: principles and practice. Nat Rev Genet 13, 303–314.

Yang, Z., and Roberts, D. (1995). On the use of nucleic acid sequences to infer early branchings in the tree of life. Mol Biol Evol 12, 451–458.

Yip, C.C., Lau, S.K., Woo, P.C., Chan, K.H., and Yuen, K.Y. (2011). Complete genome sequence of a coxsackievirus A22 strain in Hong Kong reveals a natural intratypic recombination event. J Virol 85, 12098–12099.

Zablen, L.B., Kissil, M.S., Woese, C.R., and Buetow, D.E. (1975). Phylogenetic origin of the chloroplast and prokaryotic nature of its ribosomal RNA. Proc Natl Acad Sci U S A 72, 2418–2422.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. J Theor Biol 8, 357–366.

Protein & Cell