Paraphyly and (yeast) classification
Marc-André Lachance
Department of Biology, University of Western Ontario, London, Ontario N6A5B7, Canada

Yeast systematics has wholeheartedly embraced the phylogenetic approach. Central to this has been the unspoken convention that taxa at all ranks be strictly monophyletic. This can result in a proliferation of small genera and instances of nomenclatural instability, counter to the expected benefit of phylogenetic systematics. But the literature abounds with examples, at all taxonomic levels, where paraphyly is a reality that can no longer be ignored. The very concepts of Bacteria or Archaea, under the constraint of monophyly, are in peril. It is therefore desirable to effect a shift in practices that will recognize the existence of paraphyletic taxa.

‘Paraphyly is not a contagious disease.’ Stuessy & Hörandl (2014).

The cladist dogma
Joe Felsenstein (2001) tells us that ‘statistical inference of phylogenies almost didn’t happen’. But it did, and while we continue to be dazzled by the immense progress achieved in sequencing technologies and corresponding analytical methods, the Cladistic doctrine, thought by some to be extinct, has quietly infused itself into much of contemporary systematics. Universal taxon monophyly has become an unquestioned expectation. I was first made aware of this while perusing the first edition of Plant Systematics, a Phylogenetic Approach (Judd et al., 1999), which recommended abandoning the term gymnosperms, to designate gymnosperms, in favour of the expression non-angiosperm seed plants. My hope that this would be a passing fancy gradually waned as grew the conviction that I was the only one not to see the Emperor’s extraordinary garments. As a frequent reviewer and editor of yeast species description manuscripts, I began to urge authors to demonstrate reciprocal monophyly as a sine qua non validation of their species recognition concept.

Definitions
In view of the plurality of views on the matter, a rigorous definition of key terms is warranted. A taxon is monophyletic (a clade) if it consists of complete, nested sets. The often-added codicil that a monophyletic taxon must include the ancestor (and all its descendants), besides not having been advocated by Hennig (1966, pp. 64, 148), makes no practical sense, as ancestors no longer exist and are not available to be studied, delineated, described or classified. Extant living material may consist of direct descendants of the ancestor, but not the ancestor itself. Even when a fossil is on hand, one can never guarantee that it represents the actual ancestor (Hennig, 1966, p. 72). The latter is nearly impossible given the infinitesimal fraction of living individuals that undergoes fossilization. Nonetheless, inclusion of the ancestor has found its way into many textbooks and even the scholarly literature (Donoghue & Cantino, 1988; Stuessy & Hörandl, 2014). Similarly, inclusion of the ancestor is sometimes part of the definition of paraphyly. Ashlock (1971) indicated that one should specify that the ancestor is cladistically a member of the taxon (as may or may not be the case in Farris, 1974), but it appears that such a distinction has escaped the attention of many.

A taxon is usually taken to be paraphyletic if it contains many, but not all, descendants of an ancestor. This unfortunately fails to distinguish paraphyly from polyphyly, given that all species ultimately share a common ancestor. Hennig (1966, p. 146) defines the paraphyletic taxon as one whose members share ancestral character states (symplesiomorphies) as opposed to shared derived states (synapomorphies), which underlie monophyletic taxa, or shared convergent states (homoplasy), which characterize polyphyletic taxa. However, it is possible to contrive cases when a polyphyletic taxon could also be defined on the basis of symplesiomorphies. Years of personal struggle have led me finally to understand that a paraphyletic taxon (a grade) is simply a clade that is missing a single, complete subclade. This has been put most eloquently by Stephen Mount (2010): a paraphyletic taxon is the relative complement of a clade. By contrast, polyphyletic taxa combine elements from distinct subclades and exclude parts of one or more subclades. The three taxon structures are illustrated graphically in Fig. 1, taken from Hennig (1965).

In an attempt to obviate some of the difficulties arising in the application of these concepts, Ashlock (1971) proposed the term holophyletic to designate a complete clade and regarded holophyly and paraphyly as two variants of monophyly, redefined on the basis of common ancestry only. Whether paraphyly is a variant of monophyly or polyphyly
in no way changes the meaning of the term, but can shape one’s bias for or against paraphyly on less than rational grounds, and thus the distinction does not contribute usefully to resolving the issue.

Cladistics and evolutionary taxonomy

The Cladistic school of thought advocates a system in which all taxa are monophyletic and has led to the proposal of the so-called Phylocode (Cantino & de Queiroz, 2010). Although the Phylocode is unlikely to supplant the Linnaean system in the foreseeable future, an inordinate fondness for monophyly has taken hold. Considerable resistance to this shift came from Ernst Mayr (1981). Mayr’s insistence on defining ‘evolutionary significant’, instead of monophyletic, taxa met with strong opposition (Ghiselin, 1985) and is now almost dismissed as a historical footnote. But it has remained a bee in my bonnet, sometimes making me ill at ease with the pursuit of universal monophyly in my own area of yeast systematics. The unspoken justification was that our classification, and therefore our nomenclature, would stabilize, as there is only one true phylogeny. Technically speaking, this makes the incorrect assumption that we had, from the very beginning, turned simple sequences into correct phylogenies. Any phylogeny is a hypothetical inference that relies on data quality, efficacy of the reconstruction method and compliance of the true phylogeny to our evolutionary models. In fact, phylogenetic hypotheses change over time and cause an incessant realignment of the nomenclature to accommodate current views. The result is a taxonomy that is of little value to anyone but its originator.

Cronquist (1987) stated that a cladistic classification fails to recognize fundamental differences between groups of organisms such as prokaryotes and eukaryotes. He viewed the rejection of paraphyletic groups as the ‘fatal flaw’ of cladistics. Brummit (1997) presented a well-argued case for a classification that builds on the best that both the Linnaean and the Hennigian models have to offer. The main point is that both systems are diminished by excluding the qualities of the other. But others saw differently. For example, Donoghue & Cantino (1988) considered ‘paraphyletic higher taxa […] to be artificial classes that should not be recognized formally because they are likely to confound the study of evolution’. Donoghue (1985) even rejected the biological species concept precisely because species are not necessarily monophyletic. Yet, if any taxonomic category has any claims to being real, what philosophers call a natural kind, it is the species.

My own area abounds with examples (Barnett, 2004) where the integration of methodologies that unravel phylogenetic relationships has caused numerous nomenclatural changes, including the rejection of a name to be followed by its later reinstatement. One such case is that of the genera Schwanniomyces and Debaryomyces, created (Klöcker, 1909) to accommodate two species that formed rugose ascospores. In the latter genus, an equatorial ledge was formed, which was viewed as ‘rather extraordinary’ by the author. In the Linnaean tradition, morphology of sexual features, in this case ascospores, was the primary criterion upon which yeast species were assigned to genera. In their very first published sequence-based phylogeny of a set of ascomycetous yeast species, Kurtzman & Robnett (1991) determined Schwanniomyces occidentalis to be closely related to Debaryomyces vanrijii, and thus that the monotypic genus Schwanniomyces appeared to be imbedded within the genus Debaryomyces. The authors immediately placed the genera into synonymy, thus averting paraphyly in Debaryomyces. Two decades later, all known ascomycetous yeast species had been studied by phylogenetic sequence analysis. Debaryomyces as understood at that point (Kurtzman & Suzuki, 2010) had again become paraphyletic because some Debaryomyces species appeared more closely related to species of the Kurtzmaniiella clade. Moreover, an outlier species (Debaryomyces carsonii) joined another clade that was paraphyletic with respect to all other Debaryomyces species. The authors solved the problem by reassigning some, but not all, species concerned to five separate genera, including a larger, revitalized Schwanniomyces. Some intermediary (Candida) species were set aside, presumably for a later renaming exercise. A good portion of this confusing state of things could have been averted by recognizing paraphyly in Debaryomyces. Similar rearrangements are now at risk of taking place in a clade that contains the medically important species Candida albicans because traditional criteria pertaining to sexual life cycles and other growth characteristics do not always follow the same lines as sequence-based phylogenies. One hopes

Fig. 1. Monophyly, polyphyly and paraphyly, as illustrated by Hennig (1965), reprinted with permission from the publisher.
that specialists will proceed with the utmost circumspection when dealing with conflicting evidence. In some cases, this may involve retention of well-characterized genera in spite of paraphyly, in preference to the creation of several small, indistinguishable genera.

To what extent nomenclatural changes in yeast systematics have been in conscious reaction to the detection of paraphyly is difficult to establish, as the term seldom appeared within any sizeable group containing many polytypic terminal taxa. Much to our astonishment, we found this to be the case in this study.

To what extent nomenclatural changes in yeast systematics have been in conscious reaction to the detection of paraphyly is difficult to establish, as the term seldom appeared within any sizeable group containing many polytypic terminal taxa. Much to our astonishment, we found this to be the case in this study.

**Monophyly is often an artefact of phylogenetic reconstruction**

Phylogenetic trees picture the evolution of species as a series of symmetrical bifurcations, each corresponding to a presumed allopatric speciation event. In a world designed to please practitioners of phylogenetics, all speciation would be allopatric; but it isn’t. I came to this realization while pondering the phylogeny of a novel yeast species that appeared to be an intermediary step in a cascade of two peripatric speciation events (Lachance & Fedor, 2014). Two progeny species formed sister clades with some, but not all, members of their progenitor species. In such cases, the terms mother and daughter species are more appropriate than the usual sister species. A search of the relevant literature made me aware that such a phenomenon is anything but exceptional. A special term, parasppecies, was even coined by Ackery & Vane-Wright (1984) to designate grade species that continue to share a common gene pool after the peripatric removal of a cladospecies. In their study of milkweed butterflies, they concluded: ‘[We] intuitively predicted that we might to find equal numbers of clado- and para-species within any sizeable group containing many polytypic terminal taxa. Much to our astonishment, we found this to be the case in this study.’ Albert et al. (2011) reviewed instances of parasppecies in fish and other animals, concluding that they are infrequent. However, Riesberg & Brouillet (1994) reached the opposite conclusion with respect to plants. In their estimation, most speciation begins with isolation of a peripheral population, a view that interestingly was held also by Hennig (1966, pp. 58–60). The mother species remains paraphyletic long enough for symplesiomorphies to diverge from traits of the daughter species and acquire synapomorphic properties, which in the case of plants, can be a considerable amount of time. Donoghue (1985) preferred the term metaspecies for the residual progenitor species, arguing that it is difficult to determine whether the progenitor is in fact mono-, para- or polyphyletic. Crisp & Chandler (1996) estimated that up to 50 % of plant species are paraspecies or metaspecies. Funk & Omland (2003), in an excellent conceptual review paper, reported paraphyly in 23 % of many animal species examined. That a typical speciation event gives rise to a smaller, monophyletic species carved out of a larger mono- or paraphyletic (stem) species makes sense (Stuessy & Hörandl, 2014), and the more one thinks about it, the better the chances that the stem group will in fact be paraphyletic.

Species are the unit of evolution, but repeated speciation gives rise to higher taxa. The broader question is then whether genera or families, like species, follow asymmetrical phylogenetic paths. Orthodox Cladists (Hennig, 1966; Wiley & Lieberman, 2011) insist that the unique event of phylogeny is cladogenesis, i.e. new taxa arise from the binary splitting of ancestral taxa with the concomitant disintegration of the latter. This frame of mind may account for the widespread neglect of ancestral taxa. Woese & Fox’s discovery (1977) of three instead of two major clusters of ribosomal gene divergence resulted in the reassignment of certain prokaryotes to a separate urkingdom. Woese’s tree (Woese et al., 1990), represented in Fig. 2, like other iconic trees of life (e.g. Pace, 1997), show three well-resolved clades. Arguments have been exchanged over the exact position of the root along one of the three naked stems leading to monophyletic eukaryotes, monophyletic archaeaeans, and monophyletic bacteria. Discussion of the root position diverted attention away from the unresolved and potentially embarrassing nature of the naked stems. The concept of phylogenetic dark matter comes to mind. A recent report by Spang et al. (2015) shines some light into that darkness. Environmental sequences provided evidence for a novel group of Archaea, the Lokiarchaeota, a probable sister taxon to the eukaryotes, rendering the Archaea themselves paraphyletic, as suggested earlier by Williams et al. (2013). The tree of life must therefore be redrawn with the eukaryotes emerging from within the Archaea. Marriott & Allers (2016) did exactly that (Fig. 3), but strangely concluded that the recent discoveries force us to abandon the three-domain model in favour of a two-domain model... consisting of three distinct domains. Lo the power of monophyly! Weaker evidence places the Archaea themselves in proximity to Gram-positive bacteria (Gupta, 2000; Cavalier-Smith, 2002; Valas & Bourne 2009), such that the kingdom Bacteria may also be regarded as paraphyletic, and on that basis, the two-domain tree should be reduced further to a single-domain tree (but still comprising three domains).

Paraphyly is widespread. Consider vertebrate diversification. Following the symmetrical dichotomous model of phylogenetic diversification, amphibians and fish would share a common ancestor that was neither fish nor amphibian; this cannot be. The evidence is that amphibians arose in the Devonian from a fish population that began the process of differentiating fins into feet. The progenitor species was probably a lungfish (Austin et al., 2015). Up to that
very point, the craniates consisted entirely of fish, namely the lampreys and the hagfish, the sharks and the rays, the ray-finned fish, and the lobefin fishes. Fish were fish. A hundred million years of being fish came to an abrupt end, where fish as a natural kind were sacrificed at the altar of monophyly. Similarly, appearance of the first shark rendered the progenitor agnathan paraphyletic, the birth of the first ray-finned fish from a population of a shark-like species rendered the Chondrichthyes paraphyletic, and so on with the lobefins.

Were we to persist in a system that demands universal monophyly, we shall have no choice but to do away with the Bacteria and the Archaea, elevating each of their constituent phyla to the rank of kingdom, and likewise for orders and families of fungi, plants, animals and other eukaryotes. The current solution to these problems is either to ignore the fact that many important taxa are paraphyletic or to subdivide them into innumerable monophyletic taxa as needed to eliminate paraphyly – until those very groups are found also to be paraphyletic. As stated by Brummit (1997), it ‘does not matter how many genera are recognized, genus Y must have evolved from within a genus other than itself, which is still paraphyletic. Indeed, the newly recognized genera must also have had their origins outside themselves, and increasing the number of monophyletic genera merely increases the number of potential paraphyletic genera. One can never get rid of paraphyly, though in practice it may not be apparent from a cladogram because the ancestors are not included in the analysis […]’. For some, the solution is to abolish the use of taxonomic ranks altogether. I do not see it necessary to expand on that view; interested readers should read Giribet et al. (2016).

---

**Fig. 2.** Woese’s Tree of Life, modified to show three ill-defined stem groups, Bacterial Dark Matter (BDM), Archaeal Dark Matter (ADM), and Eukaryotic Dark Matter (EDM).

**Fig. 3.** Paraphyly of the Archaea (Marriott & Allers, 2016). How many domains?
Return of the lost child: paraphyly

‘If I have had any disappointment in my decade of defending paraphyletic taxa, it is that although literally hundreds of fellow taxonomists have told me that they agree with what I say, I have persuaded very few to say the same things in public.’ (Brummit, 2003).

Taxonomy sits at an ill-defined confluence of science, craft, ontology and the law. Conscientious systematists aim at making sense of the living world by organizing species into categories that are meaningful not only to themselves, but also to other biologists and the rest of humanity. The last decades have seen undue emphasis placed on creating categories that reflect an erroneous model of cladogenesis, one that is strictly dichotomous and symmetrical. A large proportion of taxa arise not by fission of an ancestral progenitor, but by a budding process that leaves the progenitor enriched by the presence of its offspring. The retention of cohesive parent taxa under a single designation, at the same rank as progeny taxa, has the potential to prevent disastrous nomenclatural practices.

In practice, I am not suggesting that existing taxbe reshuffled and renamed to fit a peripatric model, resulting in even more instability. Every effort should be made, in delineating species and taxa of higher ranks, to demonstrate reciprocal monophyly. I strongly advocate, however, the retention of existing paraphyletic taxa where the evidence clearly points in that direction and where the alternative would be the creation of numerous, small and meaningless taxa, purely for the sake of inferred monophyly.

Hörandl (2006) made a strong case for the recognition of paraphyletic taxa and a synthetic approach to classification that integrates cladistics and evolutionary systematics. Hörandl & Stuessy (2010) further proposed a five-step course of action: (1) construct a sequence-based phylogeny; (2) map phenotypic characters and (3) compare the degree of divergence at the genetic and phenotypical levels; (4) delineate taxa on the basis of both descent and divergence, pointing at paraphyly, when present, as a feature of a taxon but not as a criterion used in the delineation; and (5) choose an appropriate rank. All else being equal, this approach can be used to erect a taxonomy whose nomenclature is stable enough to be useful.

Carter et al. (2015) proposed Paracladistics as a reformed approach to Cladistics that remains compatible with the Phylocode but accepts the approach to Cladistics that remains compatible with the of phylogenies. I have argued (Lachance et al., 2016) that a stable phylogeny is the best we can hope for because we can never guarantee that any phylogeny is correct. A stable phylogeny can arise from a broad sampling of whole genome sequence data. But even with the best data and the best methods, the continued discovery of new taxa has the potential to alter phylogenies significantly.

Prospective

In 1860, Darwin struggled over the phylogenetic relationship between marsupials and placentals (Archibald, 2012). He drew two trees, one showing the split of a hypothetical ancestor that was neither marsupial nor placentals into two modern subclasses, and another depicting ancient marsupials as the stem species from which emerged the placentals and modern marsupials. It is not clear which hypothesis Darwin favoured at the time, but it appears that he anticipated the current conundrum.

Yeast systematists generally avoid categorizing themselves as evolutionary taxonomists, Hennigians, transformed cladists or anything else of that sort. The community has an inner core of pragmatists who have earned respect by combining high productivity, seminal papers and high-quality data. These individuals act as role models to a larger, heterogeneous cohort of individuals who do not necessarily regard themselves as systematists, but contribute in various ways to the discovery of novel species and their preservation. They rarely engage in ontologico-political debate typical of what was described by Felsenstein (2001) in his short historical account. This is a good thing. However, there are times where ad hoc fashion drifts in perilous directions. In this article, my goal was to convince the community that we have shifted our emphasis excessively in favour of presumed phylogeny, with a misguided fixation on monophyly, at the detriment of our classification system as a sensible means of organizing and naming yeast taxa in a coherent, stable and useful way. Acceptance of taxa as paraphyletic when they obviously are will be one major step in our return to good sense.

Acknowledgements

I thank the Natural Sciences and Engineering Council of Canada for 36 years of continuous funding. M. Sanita Lima kindly brought Felsenstein’s essay to my attention.

References


