IUMS 2011 – The Unlimited World of Microbes – Sapporo, Japan – September 6-10 2011
JSCC and WFCC Joint Symposium on Culture collections and microbial systematics

Microbial species descriptions: the importance of multiple strains

Marc-André Lachance
Department of Biology, University of Western Ontario, London, Ontario, Canada N6A5B7
Email <lachance@uwo.ca>

Given the right cautionary preamble, most microbial systematists would probably agree that species can be defined as groups of individuals that share certain common characteristics and are distinct from other such groups. To establish groups as distinct requires an appreciation for the variance within the groups. It follows that species cannot be properly delineated if only a single representative of each species is available for study. The view presented here is that descriptions based on multiple isolates make for better science! If, for example, a sequence-based phylogenetic species concept is implied, as it often is, reciprocal monophyly must be established, which requires that the extent of sequence polymorphism be evaluated, at the very least in the new species. However, significant proportions of yeast species have been and continue to be described on the basis of a single isolate or a few isolates of dubious independence. This practice is even the object of vigorous advocacy (Kurtzman 2010). In recent years, the task of discovering new yeast species has been greatly enhanced by the availability of a continuously updated database of sequences for the D1/D2 domains of the large subunit rRNA gene. Kurtzman and Robnett (1998) have shown empirically that polymorphic species seldom exhibit more than three substitutions in that region and that well-defined species seldom differ by less than 1% substitutions. This observation has triggered a veritable avalanche of species descriptions based solely on this criterion, and often applied to single strains. Recent evidence suggests that species that are sampled thoroughly may exhibit substantial amounts of polymorphism at the level of barcoding sequences. I shall review examples of such cases and discuss their potential impact on the problem of correct delineation and typification.

Introduction

The yeast species concept has evolved extensively in recent years. Traditionally, the classification of taxa at the level of genus and above has been based on the morphology of the sexual spores and the structures involved in their production. Yeasts are diverse in this respect, as they belong to two fungal phyla, the Ascomycota and the Basidiomycota, which differ, respectively, by the formation of internal or external meiotic spores. Species assignments were initially inferred from differences in growth test responses, with the assumption was that these differences can serve as a proxy for a more objective criterion such as reproductive compatibility (biological species concept). Only a minority of known yeast species require the conjugation of distinct haploid strains (mating types) prior to entering a sexual cycle (haplontic hetero- thallism). Instead, many yeast species are either diplontic, *e.g.*, *Saccharomyces cerevisiae*, haplontic and homothallic, *e.g.*, *Schizosaccharomyces pombe*, or simply asexual, as in the large form-genera *Candida* and *Cryptococcus*. These modes of reproduction preclude the use of mating compatibility as a criterion for species assignment. DNA/DNA reassociation later served as a genetically based substitute with much success, but was often seen as onerous. DNA sequencing has now displaced previous approaches to a large extent, although a clear theoretical framework underlying sequence-based species delineation is still lacking. The main interpretation of DNA sequence data in yeast species delineation derives from the empirical observation, by Kurtzman & Robnett (1998), that strains considered conspecific by other methods rarely differed by more than three substitutions in the D1/D2 domains of the large subunit nuclear rRNA gene; strains of distinct species usually differed by at least 1% substitutions. This came to be interpreted by many as a hard and fast rule, generating a deluge of descriptions based solely on this criterion and often applied to a single isolate. The number of yeast species recognized in the five editions of "The Yeasts, a Taxonomic Study" is shown in Fig 1. The well over two-fold increase observed between 1998 and 2011 is attributable mainly to the ease with which the sequence divergence criterion can be applied to large yeast collections.

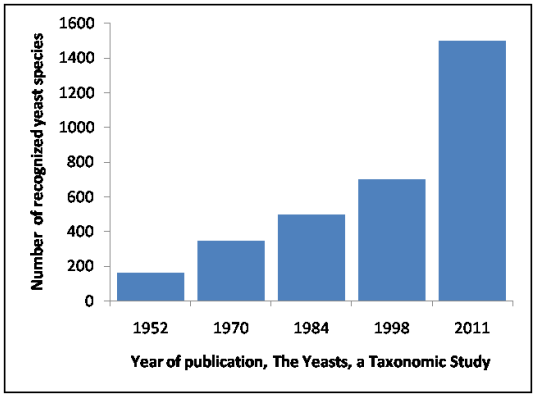

Fig 1. Increase in the number of recognized yeast species since 1952 (The Yeasts, a Taxonomic Study).

The premise of the discussion to follow deviates from the ancient but lingering view (Wilkins 2009) that species are links in a continuous chain of forms (principle of continuity) and that phylogenetic trees will gradually "fill out" as sampling is intensified (principle of plenitude). Instead, two assumptions are made. First, species are real. As presciently noted by Linnaeus in his *Systema*, taxa above the species level may be the result of human invention, but species are the work of nature. Second, gradual change within species eventually leads, through the action of divergence and extinction, to disjunct species. In Darwin's (1859) lyrical prose:

> "As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications."

The importance of multiple strains

Replication is a fundamental requirement of ecological and evolutionary research, and there is no reason why systematics should be exempt from that need. Some specific arguments in favor of documenting new species from multiple isolates have been summarized by Kurtzman (2010). Multiple strains are necessary in order (1) to assess the internal diversity of a species and (2) to determine its ecological range. The author rightly cautions that multiple isolates from the same habitat often are genetically identical and contribute little to defining real variance; indeed isolate independence should be documented in species descriptions. Regrettably, he neglects to mention the risk that a strain might be a haploid mating type, which would cause its sexual cycle to remain undetected. This would go a long way towards explaining why approximately a third of described yeast species are known only from their anamorph (asexual stage). Kurtzman (2010) argues in favor of single-strain species descriptions by invoking (1) the potential discovery of new germplasm of importance in science and biotechnology, and a better understanding of (2) biodiversity and (3) phylogeny. To these he adds the historical fact that (4) one third of known yeast species were founded on single isolates and so that valuable information would have been lost, had authors awaited the discovery of additional strains. This last observation is hardly an argument at all. First, no one is suggesting that previously described species should retroactively be stripped of their status. Second, one can counter, *ad absurdum*, that because a large number of yeast species were described without recourse to DNA sequencing, sequence data should not be required in future descriptions. The notion that our knowledge of useful microorganisms or biodiversity (1 & 2) might be hindered by delays in the description of species may be a real problem, although Kurtzman (2010) himself provides the solution, namely that the virtues of unassigned strains can be communicated in the literature and their sequences added to databases. He also points out that deposition of unnamed cultures in collections is not always common practice; again, this is easily addressed.

The rest of this discussion will focus on Kurtzman's (2010) third claim, that single-strain species descriptions positively contribute to our understanding of phylogeny. The issue has two sides. One is the easily dismissed implication that a unique lineage must be given a species name in order to be included in a phylogenetic tree. The second aspect, however, is significant and deserves some elaboration.

The phylogenetic species concept

Suh *et al*. (2004) regarded Kurtzman & Robnett's (1998) sequence divergence criterion as a phenetic species concept, advocating instead a phylogenetic concept based on reciprocal monophyly (and noting that the two models are usually in agreement). In a systematic review of fungal species recognition, Taylor *et al*. (2000) favored phylogenetic species concepts over others because of their broad applicability. They promoted the use of phylogenetic congruence (genealogical concordance) among independent loci as a means of identifying species boundaries, as recombination within a species causes gene trees not to be superimposable. This view is consistent with Hennig's (1966) distinction between tokogenetic (parent-offspring) relationships, observed among members of a species, and phylogenetic relationships, which exist between species. Kurtzman (2010) embraced the "Genealogical Concordance Phylogenetic Species Recognition" (GCPSR) of Taylor *et al*. (2000) and gave examples where multi-locus analyses provided evidence of incongruence in some yeast species. He suggested that discrepant trees can be taken as evidence for hybridization and further that multi-locus analyses are capable of detecting polymorphisms, even in the absence of multiple isolates. A clear justification for either claim was lacking, however. Demonstration that a strain is a hybrid should require that the strain be shown to contain significant portions of the genomes of two distinct species. The evidence provided by Kurtzman (2010) was the presence, in one species, of a gene whose sequence was nearly identical to that of a sister species. This could also be attributed to other causes, for example an anomalous

rate of divergence. More perplexing was the claim that multi-locus analyses can be used to identify polymorphisms in a sample of one. One can only hope that this arose from a non-standard definition of the term polymorphism.

A clear notion of polymorphism is vital to understanding of what a species is. Like any other phylogenetic species concept, GCPSR requires the establishment of reciprocal monophyly, which is hardly conceivable with only one individual. Reciprocal monophyly arises when two taxa take distinct evolutionary paths, causing tokogenetic variation (polymorphism) to give way to phylogenetic variation (divergence). Again, citing Darwin (1859):

"The only distinction between species and well-marked varieties is, that the latter are known, or believed, to be connected at the present day by intermediate gradations, whereas species were formerly thus connected.''

The nature of the gradations observed in a collection of strains can be evaluated by haplotype network analysis. The program TCS (Clement *et al*. 2007) constructs networks from DNA sequences and performs a statistical parsimony test for network membership based on the probability that each step in the network represents a single substitution. In other words, the analysis serves to distinguish the alleles of polymorphic loci within a single species from the disjoint, divergent orthologs of separate species. In a meta-analysis of hundreds of studies dealing with animal and plant taxa, Hart and Sunday (2007) showed that exclusive parsimony networks based on TCS analyses of barcoding sequences generally correspond to well-differentiated species as defined by other criteria such as reproductive isolation. In my laboratory, we have applied haplotype network analysis to ITS-D1/D2 LSU rDNA sequences in four cases where variation occurred in these sequences. For two asexual species complexes centered on *Candida azyma* and *Candida apicola*, we showed that the species themselves, stripped of non-members, remained polymorphic (Lachance *et al*. 2010). Moreover, each formed a network where all neighboring pairs fit Kurtzman and Robnett's (1998) species recognition criterion (three or fewer substitutions in the D1/D2 region), even though members at the extremities of the network did not. We also examined two sexually reproducing species, *Metschnikowia agaves* and *Starmerella bombicola*, where the ability to mate and to form mature ascospores can be used as a proxy for species membership (Lachance *et al*. 2011). Again, all members of the networks associated with these species were linked to neighbors by three or fewer D1/D2 substitutions, but could differ from peripheral relatives by as many as seven substitutions. In such a case, inadequate sampling could easily lead to the erroneous description of separate species.

Concluding remarks

It is unquestionable that species descriptions based on multiple strains are of better scientific quality and should be favored over single-strain descriptions. Biodiversity surveys that consist of random collections conducted for the sole purpose of discovering individual sequence variants that are then described as "novel species" should give way to well-planned ecological studies where communities are linked to their habitats. The problem has been raised for bacteria by Christensen *et al*. (2001) and acknowledged by Stackebrandt *et al*. (2002), although the practice of single-strain species description continues unabated. According to Felis & Dellaglio (2007), not only are the majority of bacterial species descriptions based on a single isolate, but the proportion of such descriptions continues to increase. The authors proposed the status of *species proponenda* for published descriptions of single isolates that appear to represent new species, which would solve the matter of typification and priority. Yeast systematists might benefit from a similar proposal, although it appears that authors in the field are increasingly aware of the importance of biological replicates, due in part to the activism of some members of that community. In the end, Kurtzman (2010) conceded that single-strain descriptions should be reserved for cases where "phylogenetic analysis shows the species to be well separated from neighboring species or if physiological and genetic characterization demonstrate novel properties", and discouraged "if the new taxon is scarcely resolved from members of a heavily populated clade." I agree wholeheartedly.

Acknowledgements

I thank the Natural Sciences and Engineering Research Council of Canada for research funding. Support from the Society for General Microbiology and the Japan Society for Culture Collections is also gratefully acknowledged.

References

Felis, G.E. & F. Dellaglio (2007) On species descriptions based on a single strain: proposal to introduce the status species proponenda (sp. pr.). Int. J. Syst. Evol. Microbiol. 57: 2185–2187.

Christensen, H., M. Bisgaard, W. Frederiksen, R. Mutters, P. Kuhnert & J.E. Olsen (2001) Is characterization of a single isolate sufficient for valid publication of a new genus or species? Proposal to modify Recommendation 30b of the

Bacteriological Code (1990 Revision). Int. J. Syst. Evol. Microbiol. 51: 2221–2225.

Clement, M., D. Posada & K Crandall K (2000) TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9: 1657–1660.

Darwin, C. (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life, 1st edn. John Murray, London, 502 pp.

Felis, G.E. & F. Dellaglio (2007) On species descriptions based on a single strain: proposal to introduce the status species proponenda (sp. pr.). Int. J. Syst. Evol. Microbiol. 57: 2185–2187.

Hart, M. & J. Sunday (2007) Things fall apart: biological species form unconnected parsimony networks. Biol. Lett. 3: 509–512.

Hennig, W. (1966) Phylogenetic Systematics. University of Illinois Press, Urbana, 280 pp.

Kurtzman C.P. (2010) Description of new yeast species – is one strain enough? Bull. BISMiS 1: 17-24.

Kurtzman C.P. & C.J. Robnett CJ (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. Antonie van Leeuwenhoek 73: 331–371.

Lachance, M.A., J. Dobson, D.N. Wijayanayaka & A.M.E. Smith (2010) The use of parsimony network analysis for the formal delineation of phylogenetic species of yeasts: *Candida apicola*, Candi*da azyma*, and *Candida parazyma* sp. nov., cosmopolitan yeasts associated with floricolous insects. Antonie van Leeuwenhoek 97: 155-170.

Lachance, M.A., T.M. Wijayanayaka, J.D. Bundus & D.N. Wijayanayaka (2011) Ribosomal DNA sequence polymorphism and the delineation of two ascosporic yeast species, *Metschnikowia agaves* and *Starmerella bombicola*. FEMS Yeast Res 11:324-333.

Stackebrandt, E., W. Frederiksen, G.M. Garrity, P.A.D. Grimont, P.Kämpfer, M.C.J. Maiden, X. Nesme, R. Rosselló-Mora, J. Swings, H.G. Trüper, L. Vauterin, A.C. Ward & W.B. Whitman (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol. Microbiol. 52: 1043–1047.

Suh, S.O., C.M. Gibson & M. Blackwell (2004) *Metschnikowia chrysoperlae* sp. nov., *Candida picachoensis* sp. nov. and *Candida pimensis* sp. nov., isolated from the green lacewings *Chrysoperla comanche* and *Chrysoperla carnea* (Neuroptera: Chrysopidae). Int. J. Syst. Evol. Microbiol. 54: 1883–1890.

Wilkins, J.S. (2009) Species, a history of the idea. University of California Press, Berkeley, 320 pp.