Stability of classification of filamentous fungi under changes in character coding strategy

P.D. Bridge & M.J. Sackin

1 International Mycological Institute, Ferry Lane, Kew, Surrey, TW9, 3AF; 2 Department of Microbiology, Leicester University, PO Box 138, Medical Sciences Building, University Road, Leicester LE1 9HN, UK

Received 16 May 1990; accepted 27 July 1990

Key words: Phoma, Penicillium, numerical taxonomy, character coding

Abstract

The characteristics of a number of filamentous fungal cultures were obtained from two previously published numerical taxonomic studies on Penicillium and Phoma. The coding strategies for some of the physiological and morphological properties employed in the original studies were re-examined and the data was re-coded by combining sets of characters into single ordered multistate characters. The different coding procedures were compared by generating average linkage (UPGMA) dendrograms which were in turn compared by calculating correlation coefficients between the final similarity matrices implied by these dendrograms. The character conversions had no significant effect on the final outcome of the clustering.

Introduction

Taxonomic characters as employed in numerical studies have often been coded as either binary or semi-quantitative, with features being recorded, respectively, as present/absent or as negative, weak or positive. With much bacterial and yeast data these characters are often single enzymes or the results of well documented pathways, and so the independence of such characters, if not certain, can often be empirically assumed [see 1, 17]. However, in practical fungal taxonomy a significant number of characters may be derived from quantitative data, such as colony diameters and conidial sizes or from a single feature such as the conidiophore structure [see 6, 15, 18]. In these cases, although a single feature may be identified, this may be the expression of a long pathway involving many enzymes. The independence of characters cannot always be assumed, for example if conidial secission is schizolytic then it cannot be rhexolytic [7]. This situation may be further complicated when characters relating to biochemistry and physiology are included with morphological features, for example single fungal metabolites may also be apparent as distinct pigments.

Numerical taxonomic studies with filamentous fungi have treated the above situations in different ways. In early studies data was in general reduced to a binary form in order to calculate similarities. This led to the generation of a number of characters for a single attribute such as five different characters for colony diameter, and sometimes...
the inclusion of a ‘no comparison’ state [e.g., 9, 10, 19]. More recent approaches have included the use of ordered quantitative characters for measurement data and nutritional tests [e.g., 2, 4, 11], and the generation of an ‘all present’ scheme, whereby morphological features such as colony colour or conidial shape are described from a set of non-exclusive characters (i.e. where grey/black colonies are coded as positive for grey and for black pigmentation separately), or where two different spore shape characters are used in order to describe predominant and minority populations in the same sample [see 14]. One of the fundamental neo-Adansonian principles employed in numerical taxonomy is that ‘every character is of equal weight in creating natural taxa’ [17], and obviously the inclusion of several separate characters for a single feature will increase the weighting on that feature [1, 17].

The implementation of a weighting system in numerical comparisons of filamentous fungi has been considered previously, and one alternative suggested was to consider features as either primary or secondary characters [e.g. 8]. In addition, the effect of scaling ordered quantitative characters by both Gower’s and Rajski’s coefficients has been considered [4]. However, very little work has been undertaken on the true quantitative effects of character weighting, most studies having relied on intuitive interpretation.

The use of taxonomic data from different study areas in the fungi, such as secondary metabolite data [e.g. 5], physiological and biochemical activities [e.g. 2, 3] and ultra-structural studies [e.g. 12], in addition to morphological studies has become so extensive that it is necessary to compare and/or combine such data for an integrated taxonomy [e.g. 2, 4]. This paper investigates the effects of coding strategy on sample data from such studies and uses a quantitative measure of comparison.

**Materials and methods**

*Data used.* All of the physiological, biochemical and morphological data was taken from published numerical taxonomic studies for *Penicillium* (100 characters; Table 1) and *Phoma* (92 characters; Table 2) [2, 13]. The *Penicillium* strains used were; *Penicillium expansum* IMI 285521, 232297, 28619, 297898, 297899 & 174158 *P. aurantiogriseum* IMI 89372, 285514, 40236ii, 297971, 291194, 159109 & 291190, *P. echinulatum* IMI 285508, 68236, 68236ii, 281776 & 40028. The *Phoma* strains used and the clusters they were assigned to by Monte et al., [13] were: Cluster A1 *Ph. cava* IMI 184366 &163515 and *Ph. leveillei* 331917; Cluster A2 *Ph. fimeti* IMI 163514, 331908 & 331910, *Ph. cava* 331911; Cluster B *Ph. exigua* var. *exigua* IMI 173144, 299339 & 331918, *Ph. exigua* var. *inoxydabilis* IMI 194763, *Ph. exigua* var. *linicola* IMI 197074, *Ph. exigua* var. *diversispora* IMI 331907, *Ph. medicaiginis* var. *pinodella* IMI 299428: Cluster E *Ph. epicoccina* IMI 178513 and Cluster F *Ph. sorghina* IMI 300771, 303698, 307849 & 303696.

*Character coding.* The data were initially used in the forms detailed in the original publications [2, 13]. The data was then reexamined for alternative coding strategies. In the original studies, characters derived from common features were coded as sets of separate features. Combining characters is the opposite extreme to this and so will be a severe test for the stability of the classifications. For the *Penicillium* data (see Fig. 1) the 100 characters were reduced to 82 by (1) combining growth and basic reaction on carbon sources into single multistate characters, (2) combining the three branching level characters into a single character, and (3) by combining the five conidial ornamentation characters into a single quantitative character ranging from nearly smooth to echinulate. For the *Phoma* data (see Fig. 2) the 47 morphological characters were reduced to 34 by (1) combining the nine colony size, conidial size and conidial length/breadth ratio characters into three single quantitative characters, (2) combining the grey-olivaceous and black-olivaceous pigmentation characters into single characters for both colony colour and colony reverse, (3) combining colony reverse black-green and black-green to glaucous, (4) combining the flat and high