Polish Journal of Microbiology
2012, Vol. 61, No 3, 169–181

ORGINAL PAPER

Concept for Development of Expert Computer Program for Identification of Filamentous Fungi

BEATA GUTAROWSKA[1] and MAREK IZDEBSKI[1]*

[1] Institute of Fermentation Technology and Microbiology, Technical University of Łódź, Poland
[2]Institute of Physics, Technical University of Łódź, Poland

Received 10 April 2012, revised 13 June 2012, accepted 4 July 2012

Abstract

An expert program has been developed for users working in industrial laboratories who are not experts in the identification of filamentous fungi. The database of morphological growth features currently contains 12 species from the genera *Aspergillus* and *Penicillium* grown under standard conditions. The identification algorithm implemented in the database takes into account the reliability of users, which can vary over a wide range depending on the identification feature. The reliability of users was estimated on the basis of a questionnaire survey conducted among 27 non-experts, as the likelihood of a response consistent with the assessment of experts. The program works through comparative analysis of features of the fungus being identified with the expert-developed database and selection of the most likely species among the species represented by reference strains. The expert program reduces subjective mistakes and may be extended to include further fungal species and genera; it can also be supplemented with chemotaxonomic, genetic and other data.

Key words: *Aspergillus* sp., *Penicillium* sp., filamentous fungi, expert program, morphological identification

Introduction

Advanced methods for the identification of filamentous fungi are based on PCR (Edwards *et al.*, 2002), analysis of secondary metabolites (Smedsgaard and Frisvad, 1996), and analysis of chemical substances synthesized by fungi, such as cellular fatty acids (Lopes da Silva *et al.*, 1998), cell wall polysaccharides (Carbonero *et al.*, 2001) and sterols (Grandmougin-Ferjani *et al.*, 1999). However, these procedures require specialized laboratories with necessary equipment and qualified experts; furthermore, they are expensive and time-consuming.

Macro- and microscopic morphological analysis of fungi cultured in microbiological media using species identification keys is the most widespread identification method in industrial laboratories (Samson *et al.*, 2000; Frisvad and Samson, 2004). Nevertheless it is fraught with many disadvantages, as it is estimated that 50% of filamentous fungi are incorrectly identified using the method (Flannigan *et al.*, 2001). This is due to inconsistent morphological descriptions provided by authors, lack of standard culture conditions in the keys, and primarily the identification of species by unqualified staff. The issue of incorrect identification by the culture method most commonly arises with species from the genera *Aspergillus* and *Penicillium,* which are widespread in various environments. There exist databases on the morphology of fungi (*e.g.* www.mycobank.org), but they are limited rather to storage of morphological descriptions, while the available search mechanisms do not offer any estimation of likelihoods.

There are various examples of probabilistic approaches to the identification of microorganisms (Willcox *et al.*, 1973; Bridge *et al.*, 1998), and some of these have application in computer-based identification (Kozakiewicz *et al.*, 1993; Maradona, 1994). They are based on probability matrices along with Bayes' theorem. An example of a computer-assisted probabilistic identification scheme for species of the genus *Penicillium* is PENIMAT (Kozakiewicz *et al.*, 1993).

Application of the described methods for mould diagnosis based on the evaluation of morphological features is limited to the obtaining of a yes/no answer for an identification feature in the keys. Difficulties in applying existing theories result from the following facts: morphological diagnostic features are mainly descriptive with many answer variants, or else are quantitative features; and moreover many features are mutually dependent and are of differing diagnostic significance.

* Corresponding author: M. Izdebski, Institute of Physics, Technical University of Łódź, ul. Wólczańska 219, 90-924 Łódź, Poland; fax: +48 426313639; **e-mail: izdebski@p.lodz.pl**

This article presents a concept for the development of an expert computer program for the identification of filamentous fungi based on the morphological analysis of growth. The supplied program encompasses several selected species of the genera *Aspergillus* and *Penicillium*. At present our research is in a pilot phase; the aim is to present a proposal for a new method and to develop a new algorithmic principle taking account of the diagnostic significance of multivariant descriptive and quantitative features and the likelihood of a correct answer being given by non-expert users.

Experimental

Materials and Methods

Filamentous fungi. Eight species of filamentous fungi of the genus *Aspergillus* (*Aspergillus candidus* 0417, *Aspergillus niger* 0436, *Aspergillus flavus* 0420, *Aspergillus ochraceus* 0443, *Aspergillus parasiticus* 0446, *Aspergillus oryzae* 0445, *Aspergillus terreus* 0450, *Aspergillus tamarii* 0449) and four species of the genus *Penicillium* (*Penicillium chrysogenum* 0532, *Penicillium cyclopium* (complex *Penicillium aurantiogriseum*) 0527, *Penicillium aurantiogriseum* 0526, *Penicillium expansum* 0535) were used in the study. The fungi were taken from the *ŁOCK* Collection of Pure Cultures, Technical University of Łódź, Poland.

Fungal cultures. Malt extract agar (*Oxoid*) medium was used for strain cultivation; incubation was conducted at 27°C for 7 days. Subsequently an inoculation suspension was prepared in 0.85% aqueous NaCl with a density of 4×10^6 cfu/ml (colony forming units/ml). The density was determined using a Thoma counting chamber. Filamentous fungi were cultivated using standard conditions. Each species was inoculated 10 times in a quantity of 1 loop of the inoculation suspension onto separate Czapek yeast autolysate agar CYA (*Difco*) media (Samson *et al.*, 2000) and incubated at 25–27°C for 5–7 days, at relative humidity of air 70–80%, depending on the strain, to reach a stationary phase of growth. Subsequently, macro- and microscopic growth features were determined.

Identification features. The macroscopic and microscopic identification features used in our work were selected based on the literature (Samson *et al.*, 2000; Frisvad and Samson, 2004). The variants of identification features were selected on the basis of both the literature and our own observations. The quantitative features were measured using a CX41 optical microscope with DP25 video camera (*Olympus*) and Cell-B software (*Olympus*). Our results were evaluated based on 10 replicates.

Knowledge base development, computer program. A knowledge base of four experts was created for the moulds based on morphological growth features. The data on macro- and microscopic features were entered into the database created in MS Access using appropriate forms prepared in order to facilitate communication with experts. In the case of quantitative identification features the results from a number of studies of the species were entered. Moreover, experts analysed the results of the survey conducted among non-experts and entered the resulting likelihood of the right answer. This data set allows for automatic statistical analysis carried out within the database, and the resulting significance and selectivity of each feature for a given species may be viewed. Furthermore, experts prepared explanations of the meaning of features with schemes and descriptions of their variants. A questionnaire was created for users who perform identification for entering descriptions and values obtained in measurements. The user can fill out all or only selected active boxes which describe morphological features. The form has dictionary fields prepared by experts. User form data are processed according to the rules of inference.

Mathematical Analysis Methods

Statistical analysis. The results obtained by experts were subjected to statistical analyses, namely the Pearson test was applied to reject unreliable results and the existence of normal distribution in multiple measurements was confirmed using the Kolmogorov-Smirnov test. Then calculations of mean (X) and Pearson variation coefficient (V – standard deviation to the mean X) were performed for each quantitative feature within a single strain.

A questionnaire form was prepared for the purpose of estimating the reliability of users performing identification of filamentous fungi based on macro- and microscopic features. The form contains questions about all identification features used for the description of fungi, a key with the variants of the features and colour palettes for selecting answers. The survey was conducted among 27 people not being experts in the identification of filamentous fungi. The results were compared with the assessments of experts and the likelihood P_{Rj} of the giving of the right answer for each identification feature was estimated. For the descriptive features, the likelihood that the next non-expert reply would be correct was calculated employing Laplace's law of succession:

$$P_{Rj} = \frac{X_{Rj} + 1}{X_{Nj} + 2}, \quad (1)$$

where j denotes the ordinal number of the identification feature, X_{Rj} is the number of answers consistent with the results of experts, and X_{Nj} is the number of all answers.

In the case of the quantitative features, the results of the measurement from the survey were compared with experts' results using the test, which is based on the overlap of normal distribution curves for the two sets of compared measurements

$$Z = \sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}\int_{-\infty}^{+\infty}\rho_1(x)\,\rho_2(x)\,dx = \\ = \exp\left[\frac{-(m_1-m_2)^2}{2\left(\sigma_1^2+\sigma_2^2\right)}\right], \tag{2}$$

where $\rho_1(x)$ and $\rho_2(x)$ are normal probability distributions with means m_1 and m_2 and standard derivations σ_1 and σ_2 respectively, calculated directly from the results of measurements. Equation (2) gives the result $Z=1$ whenever $m_1=m_2$. The arithmetic mean of the resulting Z_{ij} found for relevant species $i=1, 2, ..., n$ in the defined genera was taken as a measure of the likelihood of correct description of the j-th feature P_{Rj} by non-experts:

$$P_{Rj} = \frac{1}{n}\sum_{i=1}^{n} Z_{ij}\,. \tag{3}$$

The statistical significance of conformity (S_{Cij}) between the reference data provided by experts and a description of an unknown strain to identify, and an analogous significance of non-conformity (S_{Nij}), were considered for each descriptive and quantitative feature:

$$S_{Cij} = \frac{P_{Rj}}{D_{Gij}}\cdot 100\%\,, \tag{4}$$

$$S_{Nij} = \frac{1-P_{Rj}}{N_{Gj}-D_{Gij}}\cdot 100\%\,. \tag{5}$$

Here P_{Rj} is the likelihood according to equations (1) and (3), N_{Gj} is the number of tested strains in a genus for which the j-th feature has any non-empty value, and D_{Gij} is the repeatability of the value of the j-th feature for the i-th strain tested within a genus. D_{Gij} for descriptive features is defined as the number of repetitions of a species-specific feature variant within all species of a genus. In the case of quantitative features D_{Gij} was calculated as a sum of Z calculated according to equation (2) for all comparisons between the i-th strain and all species within a genus.

Rules of inference: designing the identification algorithm. The likelihood p_{ij} of conformity of data entered by the user with those in the database for the j-th feature of the i-th strain was calculated according to equation (6):

$$p_{ij} = \frac{S_{Cij}\cdot Z_{ij} + S_{Nij}\cdot(1-Z_{ij})}{\sum_k S_{Ckj}\cdot Z_{kj} + S_{Nkj}\cdot(1-Z_{kj})}\cdot 100\%\,, \tag{6}$$

where S_{Cij} and S_{Nij} are the significances according to equations (4) and (5), and Z_{ij} is the conformity of the species sought with the i-th reference strain in the database with respect to the j-th feature. In the case of descriptive features, Z_{ij} takes only the values 0 (inequality) and 1 (equality). Conformity for numerical features was determined employing equation (2) for the two sets of measurements conducted by the user and experts.

The likelihoods p_{ij} for identification features which relate to the same part of the mycelium, for example the shape, length, and width of the head, cannot be regarded as likelihoods of independent events. For each i-th strain and k-th group of dependent features, we considered the averaged likelihood $\overline{p_{ik}}$ calculated as the arithmetic mean of the respective partial likelihoods p_{ij}. Features that seem to be independent, such as filament structure, were treated as single-item groups with the likelihoods $\overline{p_{ik}}$ nequal to p_{ij} for the corresponding single features. The overall likelihood P_i for the i-th strain can be obtained by multiplying all known $\overline{p_{ik}}$ values

$$P_i = \left(\prod_{k=1}^{n_i}\overline{p_{ik}}\right)^{n_{P\max}/n_i}, \tag{7}$$

where n_i is the number of known $\overline{p_{ik}}$ for the i-th strain; n_{Pmax} is the number of known $\overline{p_{lk}}$ for the most likely l-th strain, *i.e.* the strain with the highest value of P_i. The exponent n_{Pmax}/n_i in (7) was introduced to eliminate preference for those strains with the smallest number of known product components. In the next step, the likelihoods P_i are normalized by dividing each P_i value by the sum of all P_i. The normalized likelihoods are taken as the overall likelihood of conformity P_i [%] between the data entered by the user and the i-th reference species described in the database. Calculation of P_i is a central part of the identification algorithm (see Fig. 1).

Let us define the selectivity Q_{ij} as the partial likelihood p_{ij} given by equation (6) for the case of exact correspondence of the j-th feature between an unknown strain and experts' assessment

$$Q_{ij} = p_{ij}\big|_{Z_{ij}=1}. \tag{8}$$

The selectivity is not directly used in the identification algorithm; however it appears to be a very useful parameter when considering the usefulness of the individual diagnostic features for recognition of different strains.

The calculations of D_{Gij}, S_{Cij}, S_{Nij}, p_{ij}, P_i, Q_{ij} along with the identification algorithm are implemented in an MS Access database with Visual Basic modules.

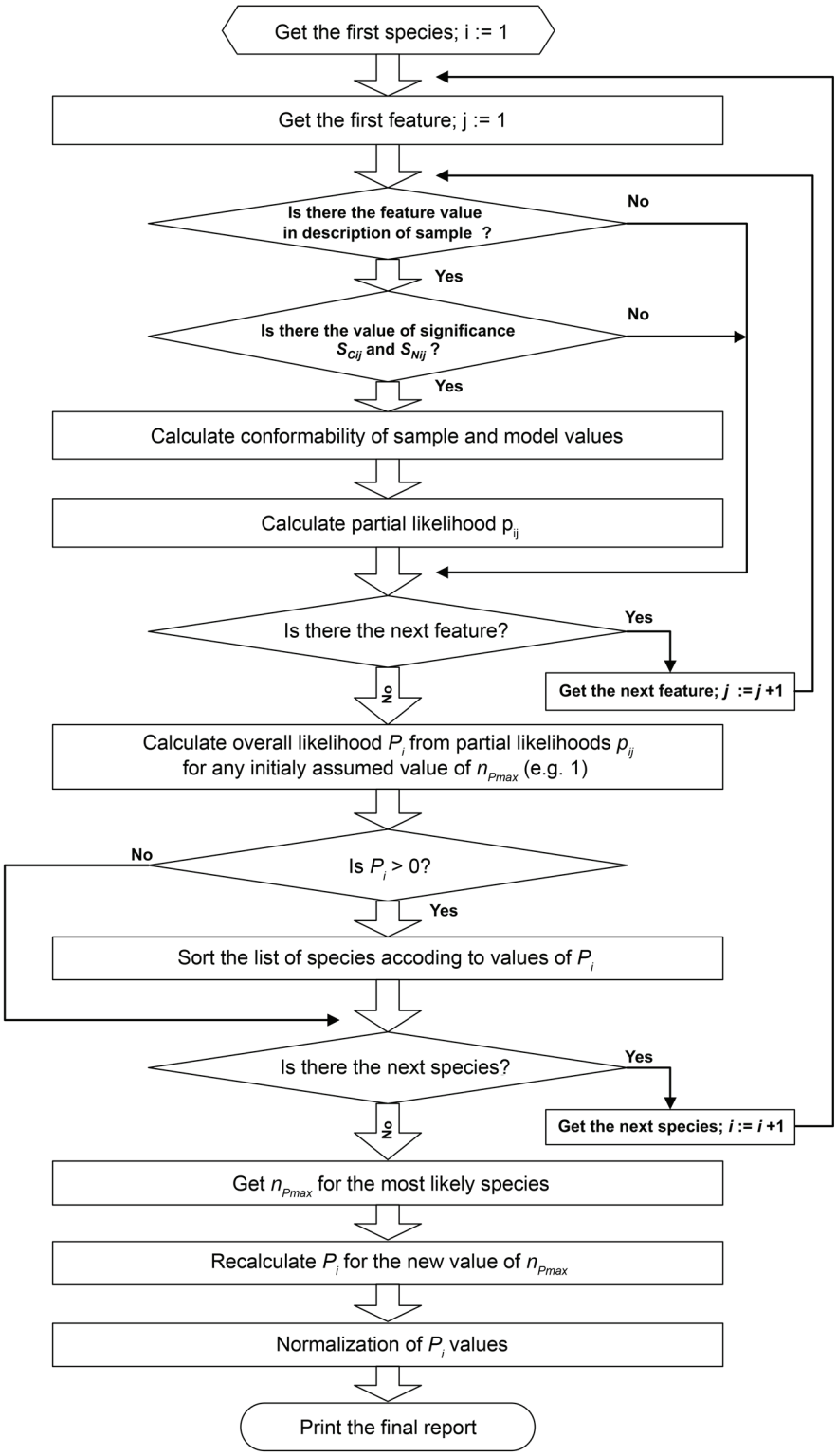

Fig. 1. Simplified algorithm for the identification of filamentous fungi species

Results

The results for the macro- and microscopic features for 12 *Aspergillus* and *Penicillium* species and the relevant statistical data are the core of the knowledge base gathered by experts (see Tables I and II, and the database: http://www.if.p.lodz.pl/marek.izdebski/expert/).

The most significant mistakes made by non-experts concern the determination of filament structure, mycelium creasing and mycelium colour for *Penicillium* (likelihood of correct answers $P_R = 0.15–0.40$). Low likelihoods of correct identification were obtained for shape of phialides and metulae ($P_R = 0.25–0.58$) (Table III).

Table I
Identification features for *Aspergillus* species on the CYA medium after 5–7 days at 25–27°C, humidity 70–80%

Feature tested	Description of identification features of the fungi from the genus *Aspergillus*							
	A. candidus	*A. flavus*	*A. niger*	*A. ochraceus*	*A. oryzae*	*A. parasiticus*	*A. tamarii*	*A. terreus*
Macroscopic features Mycelium colour	white-green	green-yellow	black	yellow	green-brown[a)], green-yellow [b)]	yellow	brown	cream-yellow
Filament structure	fluffy	granular	fluffy	granular	cotton-woolly	granular	fluffy	cotton-woolly
Occurrence of the rim	yes	no	yes	yes	no	no	yes	yes
Colour of spore-free rim	white	–	white	cream-yellow	–	–	white	white
Mycelium rim structure	regular	irregular	regular	irregular	irregular	irregular	regular	irregular
Reverse colour	yellow	yellow	yellow	brown	white	white	brown	white
Sporulation intensity	low	heavy	heavy	heavy	heavy	heavy	heavy	low
Sporulation uniformity	uneven	even	uneven	even	even	even	even	even
Mycelium creasing	yes	no	yes	no	no	no	no	no
Mycelium creasing type	concentric	–	radial	–	–	–	–	–
Mycelium creasing intensity	slight	–	deep	–	–	–	–	–
Production of secretion	no	no	no	no	no	no	no	no
Pigment secretion into the medium	no	no	no	no	no	no	no	no
Mycelium size	*X*: 30.4, *V*: 5.5	*X*: 40.3, *V*: 17.3	*X*: 43.6, *V*: 8.6	*X*: 25.1, *V*: 4.8	*X*: 42.4, *V*: 14.5	*X*: 38.5, *V*: 7.5	*X*: 48.6, *V*: 12.4	*X*: 32.8, *V*: 6.4
Mycelium height	*X*: 3.7, *V*: 22.3	nt	*X*: 3.7, *V*: 13.1	nt	*X*: 2.6, *V*: 19.9	*X*: 1.5, *V*: 35.1	*X*: 3.4, *V*: 15.2	nt
Width of spore-free rim	*X*: 2.6, *V*: 29.5	–	*X*: 4.8, *V*: 48.0	*X*: 3.2, *V*: 19.3	–	–	*X*: 3.7, *V*: 29.4	*X*: 2.4, *V*: 35.1

Table I continued

Feature tested	Description of identification features of the fungi from the genus *Aspergillus*							
	A. candidus	*A. flavus*	*A. niger*	*A. ochraceus*	*A. oryzae*	*A. parasiticus*	*A. tamarii*	*A. terreus*
Microscopic features Head shape	radial	radial	radial	radial	radial	bar-shaped	radial	radial
Vesicle shape	club-shaped	spherical	spherical	spherical	hemispherical	spherical	spherical	spherical
Metula shape	nt	nt	cigar-shaped	club-shaped	bottle-shaped	not visible	club-shaped	nt
Phialide shape	lanceolate	bottle-shaped	nt	bottle-shaped	bottle-shaped	bottle-shaped	bottle-shaped	lanceolate
Conidium shape	ellipsoidal	spherical	spherical	spherical	spherical	spherical	spherical	spherical
Conidium surface	smooth	smooth	coarse	smooth	smooth	smooth	coarse	smooth
Conidiophore width	*X*: 27.8, *V*: 30.0	*X*: 8.4, *V*: 13.7	*X*: 19.4, *V*: 41.8	*X*: 7.4, *V*: 19.3	*X*: 22.1, *V*: 33.3	*X*: 9.0, *V*: 25.2	*X*: 16.7, *V*: 35.2	*X*: 12.2, *V*: 25.0
Head length	*X*: 161.0, *V*: 16.8	*X*: 58.0, *V*: 35.6	*X*: 132.4, *V*: 36.5	*X*: 61.1, *V*: 29.2	*X*: 77.4, *V*: 29.7	*X*: 34.1, *V*: 24.8	*X*: 79.7, *V*: 28.7	*X*: 84.0, *V*: 18.0
Head width	*X*: 80.3, *V*: 17.6	*X*: 68.3, *V*: 30.3	*X*: 136.7, *V*: 32.6	*X*: 61.5, *V*: 28.3	*X*: 77.1, *V*: 32.5	*X*: 34.4, *V*: 25.5	*X*: 80.7, *V*: 19.6	*X*: 92.6, *V*: 13.7
Vesicle length	*X*: 118.4, *V*: 35.4	*X*: 29.4, *V*: 30.4	*X*: 60.6, *V*: 34.2	*X*: 25.6, *V*: 27.5	*X*: 54.3, *V*: 33.2	*X*: 18.0, *V*: 30.7	*X*: 42.9, *V*: 32.5	*X*: 41.6, *V*: 19.0
Vesicle width	*X*: 40.4, *V*: 32.1	*X*: 26.7, *V*: 15.3	*X*: 65.7, *V*: 29.4	*X*: 24.2, *V*: 26.6	*X*: 48.7, *V*: 32.1	*X*: 17.3, *V*: 32.3	*X*: 40.3, *V*: 35.5	*X*: 40.4, *V*: 18.3
Metula length	nt	nt	*X*: 13.0, *V*: 16.6	*X*: 5.8, *V*: 9.5	*X*: 9.0, *V*: 34.0	nt	*X*: 8.5, *V*: 19.2	nt
Metula width	nt	nt	*X*: 7.5, *V*: 29.9	*X*: 3.9, *V*: 17.3	*X*: 4.0, *V*: 18.8	nt	*X*: 5.0, *V*: 23.2	nt
Phialide length	*X*: 10.1, *V*: 28.1	*X*: 7.8, *V*: 17.3	nt	*X*: 10.7, *V*: 12.0	*X*: 11.2, *V*: 17.6	*X*: 11.0, *V*: 13.6	*X*: 12.9, *V*: 27.9	*X*: 12.4, *V*: 13.0
Phialide width	*X*: 3.9, *V*: 26.1	*X*: 3.5, *V*: 11.8	nt	*X*: 3.4, *V*: 9.9	*X*: 3.9, *V*: 16.2	*X*: 4.0, *V*: 12.8	*X*: 5.7, *V*: 6.8	*X*: 4.3, *V*: 16.8
Conidium length	*X*: 4.4, *V*: 6.4	*X*: 4.5, *V*: 9.4	*X*: 5.7, *V*: 46.6	*X*: 3.3, *V*: 8.2	*X*: 6.7, *V*: 13.1	*X*: 5.0, *V*: 8.4	*X*: 6.4, *V*: 10.6	*X*: 3.5, *V*: 10.3
Conidium width	*X*: 3.4, *V*: 6.0	*X*: 4.5, *V*: 9.4	*X*: 5.7, *V*: 46.6	*X*: 3.3, *V*: 8.2	*X*: 6.7, *V*: 13.1	*X*: 5.0, *V*: 8.4	*X*: 6.4, *V*: 10.6	*X*: 3.5, *V*: 10.3

X – mean in units [mm] and [μm] for macroscopic and microscopic features, respectively; *V* [%] – variation coefficient (standard deviation to the mean *X*); (–) the feature does not apply, nt – not tested,
[a] concerns the central part of the mycelium; [b] concerns the peripheral part of the mycelium

Table II
Identification features for *Penicillium* species on the CYA medium after 5–7 days at 25–27°C, humidity 70–80%

Feature tested	Description of identification features in the fungi of the genus *Penicillium*			
	complex *P. aurantiogriseum*		*P. chrysogenum*	*P. expansum*
	P. aurantiogriseum	*P. cyclopium*		
Macroscopic features Mycelium colour	gray-green	white-green[a)], gray-green[b)]	gray-green	gray-green
Filament structure	woolly	velvety	flocculent	velvety
Occurrence of the rim	yes	yes	yes	yes
Colour of spore-free rim	white	white	white	white
Mycelium rim structure	regular	regular	regular	regular
Reverse colour	yellow	yellow	yellow	orange
Sporulation intensity	heavy	heavy	heavy	heavy
Sporulation uniformity	even	uneven	even	even
Mycelium creasing	yes	yes	yes	yes
Mycelium creasing type	radial	mixed	radial	radial
Mycelium creasing intensity	deep	deep	deep	deep
Production of secretion	no	no	yes	yes
Secretion colour	–	–	yellow	colourless
Secretion transparency	–	–	transparent	transparent
Pigment secretion into the medium	yes	no	yes	yes
Mycelium size	*X*: 26.3, *V*: 11.1	*X*: 28.2, *V*: 4.0	*X*: 33.5, *V*: 20.7	*X*: 27.1, *V*: 6.4
Width of spore-free rim	*X*: 1.1, *V*: 29.4	*X*: 1.0, *V*: 27.7	*X*: 2.1, *V*: 27.3	*X*: 1.6, *V*: 31.1
Microscopic features Brush symmetry	asymmetrical	asymmetrical	asymmetrical	symmetrical
Brush density	compact	compact	loose	compact
Metula shape	cylindrical	cylindrical	cylindrical	cylindrical
Phialide shape	cigar-shaped	bottle-shaped	bottle-shaped	bullet-shaped
Conidium shape	spherical	spherical	spherical	spherical
Conidium surface	smooth	smooth	smooth	smooth
Conidiophore width	*X*: 4.3, *V*: 27.3	*X*: 4.2, *V*: 13.4	*X*: 3.3, *V*: 14.6	*X*: 3.4, *V*: 6.3
Brush length	*X*: 43.9, *V*: 25.9	*X*: 53.7, *V*: 7.1	*X*: 37.1, *V*: 18.7	*X*: 36.0, *V*: 34.2
Branch length	*X*: 19.8, *V*: 18.1	*X*: 25.2, *V*: 15.3	*X*: 14.7, *V*: 23.3	*X*: 17.1, *V*: 34.2
Branch width	*X*: 4.9, *V*: 10.3	*X*: 4.4, *V*: 10.8	*X*: 3.8, *V*: 14.3	*X*: 3.6, *V*: 10.7
Metula length	*X*: 11.9, *V*: 16.4	*X*: 15.0, *V*: 10.9	*X*: 10.1, *V*: 13.3	*X*: 9.6, *V*: 12.3
Metula width	*X*: 4.1, *V*: 17.1	*X*: 4.4, *V*: 10.7	*X*: 3.7, *V*: 14.0	*X*: 3.5, *V*: 10.5
Phialide length	*X*: 9.8, *V*: 21.5	*X*: 12.0, *V*: 16.8	*X*: 9.1, *V*: 23.0	*X*: 9.3, *V*: 22.9
Phialide width	*X*: 3.5, *V*: 14.5	*X*: 3.9, *V*: 11.0	*X*: 3.7, *V*: 20.7	*X*: 3.3, *V*: 11.9
Conidium length/width	*X*: 3.7, *V*: 6.3	*X*: 3.8, *V*: 4.1	*X*: 4.1, *V*: 10.1	*X*: 3.4, *V*: 6.8

X – mean in units [mm] and [μm] for macroscopic and microscopic features, respectively;
V [%] – variation coefficient (standard deviation to the mean *X*); (–) the feature does not apply;
[a)] concerns the central part of the mycelium; [b)] concerns the peripheral part of the mycelium

Most of the diagnostic features returned a low value for the selectivity Q_{ij}, below 50% (Table IV). Features for which higher selectivity values appeared did not however apply to all strains within the genus. Examples of such features, with selectivity $Q_{ij} > 60\%$, are mycelium colour for *A. niger* or *A. tamarii*, vesicle shape for *A. oryzae* and brush symmetry for *P. expansum*.

Our expert identification program works through comparative analysis of macro- and microscopic morphological features of the fungus being identified with

Table III
Likelihood P_{Rj} of correct identification by non-experts for identification features of fungi of the genera *Aspergillus* and *Penicillium*

Feature tested	P_{Rj}	
	Aspergillus	*Penicillium*
Macroscopic features		
Mycelium colour – central part	0.867	0.780
Mycelium colour – peripheral part	0.610	0.210
Filament structure	0.398	0.337
Occurrence of the rim	0.783	0.904
Colour of spore-free rim	0.837	0.961
Mycelium rim structure	0.765	0.916
Reverse colour	0.687	0.699
Sporulation intensity	0.928	0.831
Sporulation uniformity	0.542	0.831
Mycelium creasing	0.675	0.771
Mycelium creasing type	0.150	0.662
Mycelium creasing intensity	0.150	0.785
Production of secretion	0.940	0.410
Secretion colour	–	0.968
Secretion transparency	–	0.968
Pigment secretion into the medium	0.904	0.720
Mycelium size	0.898	0.878
Mycelium height	0.840	nt
Width of spore-free rim	0.897	0.657
Microscopic features		
Head shape	0.988	–
Vesicle shape	0.838	–
Brush symmetry	–	0.963
Brush density	–	0.825
Metula shape	0.400	0.575
Phialide shape	0.500	0.250
Conidium shape	0.785	0.725
Conidium surface	0.900	0.950
Conidiophore width	0.563	0.982
Head length	0.702	–
Head width	0.582	–
Vesicle length	0.536	–
Vesicle width	0.631	–
Brush length	–	0.955
Branch length	–	0.921
Branch width	–	0.668
Metula length	0.380	0.956
Metula width	0.777	0.605
Phialide length	0.710	0.989
Phialide width	0.666	0.805
Conidium length	0.849	0.816
Conidium width	0.734	0.642

(–) the feature does not apply, nt – not tested

the expert-developed database of fungi cultured in standard conditions (Fig. 1). Then the most likely species among the species represented by reference strains are selected. The program presents the final results in a detailed report with the values of the likelihoods P_i and p_{ij}. To evaluate the effectiveness of the method, however, it is more important to summarize the results for P_i as obtained by various non-experts (examples in Tables V and VI). High accuracy was found for identifications performed by the program for species of *Aspergillus*, with likelihoods P_i normally exceeding 99%. Some non-experts achieved lower values of P_i, although indications of the most likely strain remained accurate. In the case of the genus *Penicillium* the accuracy was significantly lower, at around 80%, when evaluation of accuracy was based on the number of identifications leading to correct indication of the most likely species. The values of P_i were nonetheless a long way from 100%, and the second-placed species often attained a double-figure percentage, sometimes up to several tens of percent.

Discussion

The most significant value of our pilot research is the development of a mathematical formula – a model for the functioning of an expert program for the identification of moulds; an innovation is the taking into account of statistical weights for individual diagnostic features and human error. An advantage of the program presented here is that it can be extended to cover other diagnostic features and that the database can be expanded to include new species. The database obtained forms the beginning of a new computer data library relating to filamentous fungi of the genera *Aspergillus* and *Penicillium*.

The results obtained here confirm previous observations of Flannigan *et al.*, 2001 on frequent identification mistakes in the fungal morphology analysis method. Low likelihoods of correct identification of fungi are due to inaccurate observations and measurements rather than feature variation. This is confirmed by the usually low values of the variation coefficient V given in Tables I and II.

It was found that the selectivities of particular diagnostic features (Table IV) are usually low within species belonging to the genera *Aspergillus* and *Penicillium*. Only selected features, important for diagnosis, take high values of Q_{ij}, although this does not apply to all species. To increase the likelihood of reliable species identification it is necessary to make use of numerous diagnostic features. The use of too few diagnostic features is the probable reason for the frequent errors in identification of filamentous fungi.

Table IV
Identification selectivity Q_{ij} [%] for identification features of fungi of the genera *Aspergillus* and *Penicillium*

Feature tested	*A. candidus*	*A. flaus*	*A. niger*	*A. ochraceus*	*A. oryzae*	*A. parasiticus*	*A. tamarii*	*A. terreus*	*P. aurantiogriseum*	*P. cyclopium*	*P. chrysogenum*	*P. expansum*
Macroscopic features												
Mycelium colour [a)]	86.2	86.2	86.2	44.2	86.2	44.2	86.2	86.2	30.5	54.2	30.5	30.5
Mycelium colour [b)]	58.8	31.7	58.8	31.7	31.7	31.7	58.8	58.8	25.0	25.0	25.0	25.0
Filament structure	13.8	13.8	13.8	13.8	17.8	13.8	13.8	17.8	27.6	21.6	27.6	21.6
Occurrence of the rim	17.1	22.8	17.1	17.1	22.8	22.8	17.1	17.1	25.0	25.0	25.0	25.0
Rim colour	23.8	–	23.8	56.2	–	–	23.8	23.8	25.0	25.0	25.0	25.0
Rim structure	22.0	16.9	22.0	16.9	16.9	16.9	22.0	16.9	25.0	25.0	25.0	25.0
Reverse colour	23.4	23.4	23.4	32.3	23.4	23.4	32.3	23.4	29.1	29.1	29.1	43.6
Sporulation intensity	40.6	16.2	16.2	16.2	16.2	16.2	16.2	40.6	25.0	25.0	25.0	25.0
Sporulation uniformity	14.1	13.0	14.1	13.0	13.0	13.0	13.0	13.0	31.2	62.1	31.2	31.2
Creasing	20.5	14.4	20.5	14.4	14.4	14.4	14.4	14.4	25.0	25.0	25.0	25.0
Creasing type	15.0	–	15.0	–	–	–	–	–	28.5	39.5	28.5	28.5
Creasing intensity	15.0	–	15.0	–	–	–	–	–	25.0	25.0	25.0	25.0
Production of secretion	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	20.5	20.5	20.5	20.5
Secretion colour	–	–	–	–	–	–	–	–	–	–	96.8	96.8
Secretion transparency	–	–	–	–	–	–	–	–	–	–	50.0	50.0
Pigment secretion into the medium	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	29.5	46.2	29.5	29.5
Mycelium size	43.8	11.8	22.4	77.7	15.2	24.5	29.6	32.4	26.0	25.4	28.1	25.3
Mycelium height	23.2	nt	28.4	nt	30.2	55.1	24.7	nt	nt	nt	nt	nt
Width of spore-free rim	21.7	–	22.7	21.0	–	–	21.7	23.1	23.2	24.6	25.7	21.6

Table IV continued

Feature tested	*A. candidus*	*A. flaus*	*A. niger*	*A. ochraceus*	*A. oryzae*	*A. parasiticus*	*A. tamarii*	*A. terreus*	*P. aurantiogriseum*	*P. cyclopium*	*P. chrysogenum*	*P. expansum*
Microscopic features												
Head shape	14.3	14.3	14.3	14.3	14.3	92.2	14.3	14.3	–	–	–	–
Vesicle shape	62.2	15.8	15.8	15.8	62.2	15.8	15.8	15.8	–	–	–	–
Brush symmetry	–	–	–	–	–	–	–	–	32.9	32.9	32.9	89.7
Brush density	–	–	–	–	–	–	–	–	31.1	31.1	61.1	31.1
Metula shape	nt	nt	36.4	23.5	36.4	36.4	23.5	nt	25.0	25.0	25.0	25.0
Phialide shape	14.3	14.3	nt	14.3	14.3	14.3	14.3	14.3	20.0	16.7	16.7	20.0
Conidium shape	34.3	13.7	13.7	13.7	13.7	13.7	13.7	13.7	25.0	25.0	25.0	25.0
Conidium surface	16.1	16.1	37.5	16.1	16.1	16.1	37.5	16.1	25.0	25.0	25.0	25.0
Conidiophore width	15.1	12.3	10.4	13.7	12.2	11.5	11.1	12.2	25.0	37.6	31.2	31.3
Head length	31.4	13.7	13.3	14.4	13.3	28.1	13.5	16.0	–	–	–	–
Head width	11.7	10.7	13.0	11.2	10.3	18.8	11.6	12.0	–	–	–	–
Vesicle length	18.7	11.3	11.8	12.3	11.4	16.5	10.8	12.7	–	–	–	–
Vesicle width	11.3	13.9	17.1	12.9	12.2	18.7	10.8	14.3	–	–	–	–
Brush length	–	–	–	–	–	–	–	–	24.0	49.2	34.4	27.7
Branch length	–	–	–	–	–	–	–	–	28.2	44.8	37.8	24.5
Branch width	–	–	–	–	–	–	–	–	28.1	23.3	23.4	27.6
Metula length	nt	nt	16.8	16.8	16.0	nt	18.5	nt	29.7	68.6	33.8	38.1
Metula width	nt	nt	31.7	27.1	26.1	nt	24.8	nt	19.1	17.8	19.0	18.4
Phialide length	11.8	15.2	nt	13.8	13.1	13.6	12.5	13.9	25.3	36.5	27.4	26.4
Phialide width	12.1	14.5	nt	15.1	14.1	14.8	25.4	13.6	23.0	22.6	22.8	21.5
Conidium length	26.0	21.9	7.1	36.3	29.8	26.6	31.0	29.7	24.5	28.3	29.7	36.1
Conidium width	20.7	26.2	7.3	22.0	24.9	26.3	25.6	19.3	21.3	23.6	23.4	24.7

(–) the feature does not apply; nt – not tested,
[a)] concerns the central part of the mycelium; [b)] concerns the peripheral part of the mycelium

Table V
Examples of identification results obtained by various non-experts for fungi of the genus *Aspergillus*

Species name (attempt)	Likelihoods P_i [%] of conformity with compared species in the database							
	A. candidus	*A. flavus*	*A. niger*	*A. ochraceus*	*A. oryzae*	*A. parasiticus*	*A. tamarii*	*A. terreus*
A. niger (1)	0.00	0.02	99.88	0.00	0.01	0.00	0.07	0.00
A. niger (2)	0.01	0.02	99.11	0.07	0.21	0.09	0.40	0.09
A. niger (3)	0.00	0.01	99.37	0.01	0.22	0.03	0.28	0.08
A. niger (4)	0.01	0.03	99.90	0.00	0.04	0.01	0.01	0.00
A. niger (5)	0.00	0.00	99.90	0.00	0.04	0.01	0.05	0.01
A. oryzae (1)	0.27	23.79	2.20	0.06	40.02	1.43	14.22	18.00
A. oryzae (2)	0.00	0.91	0.10	0.02	95.64	0.54	1.30	1.48
A. oryzae (3)	0.03	2.07	0.79	0.13	83.61	8.85	4.19	0.33
A. oryzae (4)	0.00	0.01	0.01	0.00	99.95	0.01	0.01	0.00
A. oryzae (5)	0.00	0.04	0.04	0.00	99.45	0.26	0.20	0.01
A. tamarii (1)	0.00	0.00	0.00	0.00	0.00	0.00	99.99	0.00
A. tamarii (2)	0.00	0.00	0.04	0.05	0.01	0.00	99.87	0.01
A. tamarii (3)	0.00	0.01	0.07	0.01	0.02	0.00	99.89	0.00
A. tamarii (4)	0.00	0.01	0.06	0.02	0.04	0.01	99.85	0.01
A. tamarii (5)	0.34	0.01	2.16	0.01	0.01	0.00	97.45	0.01

Table VI
Examples of identification results obtained by various non-experts for fungi of the genus *Penicillium*. Erroneous indications of the most likely strains are underlined

Species name (attempt)	Likelihoods P_i [%] of conformity with compared species in the database			
	P. aurantiogriseum	*P. cyclopium*	*P. chrysogenum*	*P. expansum*
P. aurantiogriseum (1)	52.76	20.59	17.13	9.51
P. aurantiogriseum (2)	44.93	1.43	42.54	11.10
P. aurantiogriseum (3)	16.54	<u>73.87</u>	4.13	5.46
P. aurantiogriseum (4)	46.46	0.94	41.14	11.46
P. aurantiogriseum (5)	65.93	1.63	15.36	17.08
P. chrysogenum (1)	14.87	2.07	69.43	13.63
P. chrysogenum (2)	7.87	0.16	82.41	9.56
P. chrysogenum (3)	12.96	0.60	81.45	4.98
P. chrysogenum (4)	15.56	0.11	75.93	8.40
P. chrysogenum (5)	11.20	0.12	81.90	6.77
P. expansum (1)	<u>37.55</u>	0.30	27.45	34.70
P. expansum (2)	29.27	2.32	<u>41.51</u>	26.90
P. expansum (3)	6.23	36.71	3.32	53.74
P. expansum (4)	35.43	0.45	28.30	35.82
P. expansum (5)	23.36	0.70	19.62	56.32

Because the features defined as relevant for identification are often misinterpreted by non-experts, the results for the selectivity of identification features confirm the need to develop expert programs. The expert program presented here works by creating a database and determining significance values for individual diagnostic features for each species, taking account of the likelihood of a correct answer being given by non-expert users. This program differs from other programs used for microorganism diagnosis. The literature already refers to applications of likelihoods in the identification of microorganisms, the most frequently used being Bayes' theorem (Willcox *et al.*, 1973; Tardivel and Morse, 1998; Bridge *et al.*, 1998). This approach is used in programs where a two-variant answer is obtained, and is of great importance in the identification of bacteria.

In the case of fungal identification, besides features of that type we also have many others, *e.g.* measurable morphological features and descriptive multi-variant features. Moreover, many features are mutually dependent. Consistent application of the approach presented in the works of Willcox *et al.* (1973) would require the collection of a vast data set, in order to build for each descriptive feature a reliable matrix of likelihoods linking every described species with every anticipated variant of a feature.

We remained consistent with that approach in the case of features taking only two states, although these make up only a small proportion of the identification features we used. This consistency applies to the mathematical relationships used, although we have a different interpretation of what it is that the likelihoods P_R describe.

In the diagnosis of microorganisms as described by other authors there is no justification for any model feature variants indicated unambiguously by experts for a specific species. For a given species every test result has to be considered to be correct; a test value leads only to an uneven distribution of likelihoods for particular variants, and the margin for human misrepresentation of the test result is small. Our approach to the identification of filamentous fungi is based on the reverse assumption, namely that experts are able to state unambiguously one very probable variant of a given feature for a specified species, and the source of the scattering of observation results is mainly the non-experts.

High accuracy was found for identifications performed using the program for species of the genus *Aspergillus*, but the results obtained for *Penicillium* must be considered unsatisfactory. This may be a result of the low level of differentiation of features in species of the genus *Penicillium*. This leads to many difficulties in identifying *Penicillium* strains by the traditional method using keys, and this is reflected in the larger identification error obtained using the expert program.

An undoubted limitation of our research and of the functioning of the program is the small number of studied species and strains within a species, and the fact that the present database is based exclusively on macro- and microscopic identification features of colonies of moulds of two genera. It also cannot be used in cases where a strain is morphologically immature, *i.e.* does not produce structures such as conidiophores (so-called *mycelia sterilia*). The knowledge base of fungal morphology available in our expert program forms a starting point for studies which should be extended to include further species and genera; furthermore, the database can be supplemented with chemotaxonomic and genetic data and information about the secondary metabolites and volatile compounds synthesized. Extension of the analysis to include other identification features provides the possibility of more reliable identification up to species, in a case where one of the methods is insufficient. If based on a macroscopic method a mould is not successfully diagnosed up to species, the computer program would suggest what analyses are required in order to obtain answers to the diagnostic questions posed. Chemotaxonomic and genetic identification methods can provide results of high significance. However it should be considered that only a small number of fungus species have so far been genetically diagnosed (far fewer than by the method of morphological analysis), the number of genetic primers for the identification of moulds is still inadequate, and sometimes moulds are successfully identified using the available primers only up to genus. Moreover, not all laboratories have genetic analysis techniques available, so by choosing other diagnostic methods and taking account of the weight (significance) of the considered diagnostic features it is also possible to carry out identification with a high degree of accuracy.

In the course of the research discrepancies were found in the results for, for example, the colour of the mycelium as determined by experts and non-experts, although they received the same colour palette. This feature, although it is considered significant by experts, received a lower statistical weighting in our calculations. It is desirable in the future to refine the analysis of colours and to avoid errors in their determination. The tests were carried out by three experts working a single laboratory and by 27 non-experts. It is desirable in the future to confirm the research in model conditions at several specialist laboratories at least, with the participation of more than a dozen or so experts. Our fungal identification program will be developed and supplemented with new data, and in the future could be used for the identification of fungi in industrial laboratories and also in specialized research laboratories.

Acknowledgements

We would like to thank Joanna Kozłowska, MEng. and Elżbieta Piotrowska, MEng. for their assistance in the database development.

Literature

Bridge P.D., P. Jeffries, D.R. Morse and P.R. Scott. 1998. *Information technology, plant pathology and biodiversity*. CAB International, Wallingford, Oxford.

Carbonero E.R., G.L. Sassaki, P.M. Stuelp and P.A.J. Gorin, S.M. Woranovicz-Barreira and M. Iacomini. 2001. Comparative studies of the polysaccharides isolated from lichenized fungi of the genus Cladonia: significance as chemotypes. *FEMS Micobiol. Lett.* 194: 65–69.

Edwards S.G., J. O'Callaghan and A.D.W. Dobson. 2002. PCR-based detection and quantification of mycotoxigenic fungi. *Mycol. Res.* 106: 1005–1025.

Flannigan B., R.A. Samson and J.D. Miller. 2001. *Microorganisms in home and indoor work environments: diversity, health impacts, investigation and control*. Taylor & Francis, London and New York.
Frisvad J.C. and R.A. Samson. 2004. Polyphasic taxonomy of Penicillium subgenus Penicillium. A guide to identification of food and air-borne terverticillate penicillia and their mycotoxins. *Stud. Mycol.* 49: 1–174.
Grandmougin-Ferjani A., Y. Dalpe, M.A. Hartmann, F. Laruelle and M. Sancholle. 1999. Sterol distribution in arbuscular mycorrhizal fungi. *Phytochemistry*, 50: 1027–1031.
Kozakiewicz Z., P.D. Bridge and R.R.M. Paterson. 1993. Penimat: a computerised identification scheme for terverticillate Penicillium species. *Int. Biodeterior. Biodegrad.* 32: 117–125.
Lopes da Silva T., E. de Sosa, P.T. Pereira, A.M. Ferrăo and J.C. Roseiro. 1998. Cellular fatty acid profiles for the differentiation of Penicillium species. *FEMS Micobiol. Lett.* 164: 303–310.
Maradona M.P. 1994. Integrated software for probabilistic identification of microorganisms. *Bioinformatics* 10: 71–73.
Samson R.A., E.S. Hoekstra, J.C. Frisvad and O. Filtenborg (eds). 2000. *Introduction to food- and air borne fungi*. 6th ed. Centraalbureau voor Schimmelcultures, Utrecht.
Smedsgaard J. and J.C. Frisvad. 1996. Using direct electrospray mass spectrometry in taxonomy and secondary metabolite profiling of crude fungal extracts. *J. Microbiol. Methods* 25: 5–17.
Tardivel G.M. and D.R. Morse. 1998. The role of the user in computer-based identification, pp. 247–259. In: Bridge P.D., P. Jeffries, D.R. Morse and P.R. Scott (eds). *Information technology, plant pathology and biodiversity.* CAB International, Wallingford, Oxford.
Willcox W.R., S.P. Lapage, S. Bascomb and M.A. Curtis. 1973. Identification of bacteria by computer: theory and programming. *J. Gen. Microbiol.* 77: 317–330.

Supporting information

The database mentioned in the paper can be downloaded as an MDE file (MS Access 2003 format) at: http://www.if.p.lodz.pl/marek.izdebski/expert/

The database has been tested so far in MS Access 2003 and 2007 installed under Windows XP SP2 and SP3. If you encounter a technical problem, please contact the corresponding author.