Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Microbiological Methods 61 (2005) 295–304
Journal of Microbiological Methods
www.elsevier.com/locate/jmicmeth

Automated and unbiased classification of chemical profiles from fungi using high performance liquid chromatography

Michael Edberg Hansen[a,b,*], Birgitte Andersen[b], Jørn Smedsgaard[b]

[a]*Informatics and Mathematical Modelling (IMM), Richard Petersens Plads, building 321, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*
[b]*Center for Microbial Biotechnology (CMB), BioCentrum–DTU, Søltofts Plads, building 221, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*

Received 5 January 2004; received in revised form 26 November 2004; accepted 8 December 2004

Abstract

In this paper we present a method for unbiased/unsupervised classification and identification of closely related fungi, using chemical analysis of secondary metabolite profiles created by HPLC with UV diode array detection. For two chromatographic data matrices a vector of locally aligned full spectral similarities is calculated along the retention time axis. The vector depicts the evaluating of the alikeness between two fungal extracts based upon eluted compounds and corresponding UV-absorbance spectra. For assessment of the chemotaxonomic grouping the vector is condensed to one similarity describing the overall degree of similarity between the profiles. Two sets of data were used in this study: One set was used in the method development and a second dataset used for method validation. First we developed a method for evaluating the secondary metabolite production from closely related *Penicillium* species. Then the algorithm was validated on fungal isolates belonging to the genus *Alternaria*. The results showed that the species may be segregated into taxa in full accordance with published taxonomy.

Keywords: UV; Spectrum; Aligning; Similarity; Distance; *Alternaria*; *Penicillium*; Classification; Identification

1. Introduction

Chromatograms from either gas chromatography (GC) or high performance liquid chromatography (HPLC) form the basis of one of the most powerful methods in analytical chemistry, because they enable the researchers to generate fingerprints of the many compounds found in highly complex samples. These fingerprints can then be used as very effective tool for comparison, classification, identification, or typification of samples and have found widespread use in e.g. flavor research, forensic investigations, and in chemotaxonomic characterization of microorganisms and plants (Cou-

* Corresponding author. Center for Microbial Biotechnology (CMB), BioCentrum–DTU, Søltofts Plads, building 221, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark. Tel.: +45 4525 2709; fax: +45 4588 1397.
E-mail address: meh@biocentrum.dtu.dk (M.E. Hansen).

0167-7012/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2004.12.005

net et al., 2004; Egan et al., 2003; Andersen et al., 2003). Examination and comparison of many sets of chromatographic samples, which are often large data matrices when spectroscopic detection is used[1], can exceed the limit of what is possible for manual handling in further data analyses. Instead of using the full data matrices, which include all the collected data points, data can be subjected to a substantial reduction by identification of specific peaks of interest in the chromatogram, calculation of retention index and peak area, and in the case of spectroscopic data extraction, by using the peak spectrum. In this way relevant information can be extracted (retention time, amount and UV-spectrum of the eluting components), thereby reducing the amount of information to manageable portions.

Classification and/or identification of samples based on chromatographic profiles using manual peak detection and identification is subjective and the chemotaxonomic work has often been accused of being biased and therefore unreliable. Precise sample identification depends on the experience of the researcher, so when the number of samples increases, the difficulty of identifying and extracting relevant peaks and their information becomes an overwhelming task. Hence, automated and unbiased methods for pinpointing relevant retention times (RT's) and their UV-spectra are needed.

The purpose of this study was to evaluate an analysis method that utilizes the fact that different components eluting at the same time (with close to identical RT's) have almost unique spectral profiles. By comparing co-eluting components by their UV-spectra across samples, information about the (dis)similarity between actual compounds could be examined. The similarity was evaluated as a "distance" between the observed UV-spectra. Data matrices from hyphenated chromatographic analyses were used to validate the chemotaxonomic grouping of closely related, cheese associated *Penicillium* species and closely related, plant pathogenic *Alternaria* species grown in pure cultures.

[1] e.g. Ultra-Violet.

2. Materials and methods

2.1. Dataset 1

This dataset consisted of 53 isolates of the genus *Penicillium*. The isolates are listed in Table 1 and include type cultures and isolates *P. camemberti*, *P. caseifulvum*, *P. palitans*, *P. atramentosum* and *P. commune* collected from Danish dairies. The cultures were prepared by tree point inoculation from spore solutions prepared from fresh cultures on Czapek Yeast Extract agar (CYA) and Yeast Extract Sucrose agar (YES) (Samson et al., 2000; Pitt, 1979). The cultures were allowed to grow for 7 days in the dark at 25 °C. All isolates are kept at the IBT Culture Collection at BioCentrum-DTU in Kgs. Lyngby.

Cultural extract of the *Penicillium* cultures were prepared using a slightly modified version of the plug extraction method (Smedsgaard, 1997) as follows: Tree plugs (6 mm diameter) were cut from each culture, transferred to a 2-ml vial and extracted twice: The first extraction was done with 500 ml ethyl acetate containing 0.5% formic acid ultrasonically for 45 min. The ethyl acetate extract was transferred to a clean vial and evaporated to dryness in a Rotary Vacuum Concentrator (Christ Frees Driers, USA). The tree plugs were then re-extracted ultrasonically for further 45 minutes by addition of 500 µl 2-propanol. The 2-propanol extract was likewise transferred to the vial containing the residues from the ethyl acetate extraction and evaporated to dryness. The combined extraction residues were re-dissolved in 400µl methanol ultrasonically and filtered through a 0.45µm filter (Minisart RP-4, Satorius, USA) into a clean vial before HPLC analysis.

HPLC analysis was performed on a HP-1100 system (Hewlett Packard, Germany) with a HP 1100 Diode Array Detector (DAD). Separation were done on a 4 mm id*100 mm HP Hypersil BDS C18 column with a 4mm id*4mm pre-column using a linear gradient of water–acetonitrile (both containing 500 µl formic acid per litre) going from 10% acetonitrile in water to 100% acetonitrile in 30 minutes, then 100% +acetonitrile was maintained for 5 min, before returning the gradient to starting conditions. UV-spectra were collected from 200 to 500 nm with a resolution of 2 nm at a rate of 1.25 UV-spectra per second from the DAD. Fig. 1 shows an example of a

Table 1
Penicillium and *Alternaria* species used in the study with the corresponding notations

Fungal species	Identification codes			
Penicillium atramentosum	IBT 10565	IBT 11801	IBT 13139	IBT 14762
	IBT 15294			
P. camemberti	IBT 3505	IBT 11567	IBT 11569	IBT 11755
	IBT 21600	IBT 21601	IBT 21602	IBT 21603
	IBT 21604			
P. caseifulvum	IBT 10842	IBT 14761	IBT 15151	IBT 15157
	IBT 18280	IBT 18285	IBT 18725/1	IBT 18725/2
	IBT 18727	IBT 18732	IBT 19791	IBT 19799
	IBT 19801			
P. commune	IBT 3427	IBT 6200	IBT 6367	IBT 6369
	IBT 10763	IBT 13043	IBT 13836	IBT 14135
	IBT 17106	IBT 17345	IBT 18102	IBT 18715
	IBT 21605			
P. palitans	IBT 3117	IBT 6454	IBT 6911	IBT 10715
	IBT 12082	IBT 13273	IBT 13420	IBT 13710
	IBT 14112	IBT 14741	IBT 14757	IBT 15150
	IBT 15899			
Alternaria alternata	EGS 34-016	EGS 34-039	EGS 35-193	
A. gaisen	EGS 37-1321	EGS 37-1332	EGS 39-1590	EGS 90-0391
	EGS 90-0512			
A. limoniasperae	EGS 39-185	EGS 39-187	EGS 45-080	EGS 45-100
	EGS 46-159			
A. longipes	EGS 30-033	EGS 30-034	EGS 30-048	EGS 30-051
	EGS 30-080			
A. tangelonis	EGS 41-175	EGS 45-014	EGS 45-016	EGS 45-114
	EGS 45-121			
A. turkisafria	EGS 44-159	EGS 44-166	EGS 45-056	EGS 45-057
	EGS 45-058			

HPLC data matrix obtained from *P. caseifulvum* (IBT15151).

2.2. Dataset 2

This dataset included 28 isolates of the genus *Alternaria*. The isolates are listed in Table 1 and include type cultures and isolates of *A. gaisen*, *A. limoniasperae*, *A. longipes*, *A. tangelonis* and *A. turkisafria* collected from various citrus types, pear, banana and tobacco. The cultures were prepared by tree point inoculation from spore solutions prepared from fresh cultures on Potato Carrot Agar (PCA) and Dichloran Rose Bengal Yeast Extract Sucrose agar (DRYES) (Samson et al., 2000; Simmons and Roberts, 1993). The cultures were allowed to grow for 14 days in the dark at 25 °C. All isolates are kept at the IBT Culture Collection at BioCentrum-DTU in Kgs. Lyngby.

Cultural extracts of the *Alternaria* cultures were prepared using a slightly modified version of the plug extraction method (Smedsgaard, 1997) as follows: Nine agar plugs (6 mm in diameter) were cut and placed in a 2-ml-screw-top vial. All nine plugs were extracted with 1.0 ml ethyl acetate containing 1% formic acid by sonication for 60 min. The extract was transferred to a clean 2-ml vial, evaporated to dryness in a rotary vacuum concentrator (Christ, Gefriertrocknungsanlagen GmbH, Germany), re-dissolved ultrasonically in 500 μl methanol, and filtered through a 0.45-μm filter (Minisart RP-4, Satorius, USA) into a clean 2-ml vial prior to HPLC analysis.

HPLC analysis was performed on a HP-1100 system (Hewlett Packard, Germany) with a HP 1100 Diode Array Detector (DAD). Separation was done on a 2 mm id*120 mm HP Hypersil BDS-C_{18} (3 μm particle size) cartridge column (Agilent, USA) with a 10×2 mm (i.diam) Superspher 100 RP-18 guard column (Agilent, USA) using a linear gradient of water–acetonitrile (both containing 500 μl formic acid per litre) going from 10% acetonitrile to 50%

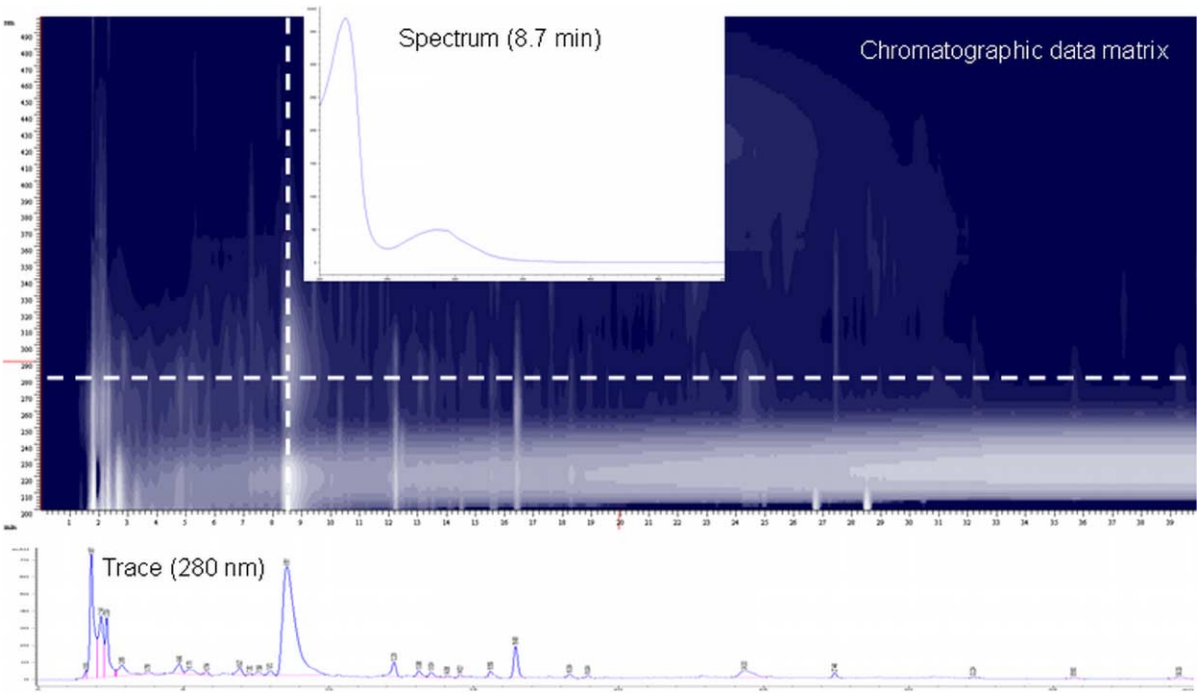

Fig. 1. Example of a high performance liquid chromatographic (HPLC) data matrix of a fungal extract from *Penicillium caseifulvum* (IBT15151). The spectrum detected after 8.7 min and the trace at 280 nm can also be seen.

acetonitrile in 30 min and from 50% to 100% acetonitrile in 10 min. One hundred percent acetonitrile was maintained for 5 min before returning the gradient to starting conditions. UV-spectra were collected from 190 to 600 nm with a resolution of 2 nm at a rate of 1.25 spectra per second from the DAD.

2.3. Calculations, equations and data analysis

The following nomenclature will be used in this paper: Chromatographic data matrices containing data of the type (time, wavelength, absorbance), are referred to as the chromatograms. The chromatographic trace at a single wavelength is called a chromatographic profile, while the set of readings for a given time point is called spectra. In equations, chromatograms and other matrices are indicated by uppercase boldface roman letters,

$$\mathbf{X}_i = \begin{bmatrix} x_{i,1}(\lambda_1) & \cdots & x_{i,1}(\lambda_L) \\ \vdots & \ddots & \vdots \\ x_{i,N}(\lambda_1) & \cdots & x_{i,N}(\lambda_L) \end{bmatrix} \quad (1)$$

where L is the length of the UV spectrum and N is the total number of spectra in the HPLC data matrix ($L \times M$). UV-spectra can therefore be regarded as vectors or points in an L-dimensional Euclidian space, in which each dimension represents the absorbance at a certain wavelength. The length of such a vector is proportional to the concentration, and vectors are represented by lowercase roman boldface letters,

$$\mathbf{x}_{i,n} = \{x_{i,n}(\lambda_1),..,x_{i,n}(\lambda_L)\} \quad (2)$$

whereas scalars are indicated by italic letters.

2.4. Similarities

The principle of method is based on evaluating the spectral information present in the HPLC data matrices. We have split up the algorithm in two parts: 1) calculation of a local similarity followed by 2) calculation of a global similarity between the HPLC data matrices.

2.5. Local similarity

Assuming the spectra in two matrices are continuous up until the order of P, it was shown that derivatives were well suitable for extracting information about shape and similarity between the UV-spectra r and s. The p'th order derivative of any spectrum x is given by

$$\mathbf{x}^{(p)} = \frac{\partial^p}{\partial \lambda^p}(\mathbf{x}^*\mathbf{h}) \tag{3}$$

where we normalize and smoothen each order or derivatives of the spectrum using

$$\tilde{\mathbf{x}}^{(p)} = \frac{1}{\max(\mathbf{x}^{(p)})}\mathbf{x}^{(p)} \tag{4}$$

and $\mathbf{h}=\{h_i\}$, $i\in\{1\ldots,N_h\}$ are filter weights eliminating small variations originating from noise fragments. Using this, a pair of UV-spectra can be evaluated based on

$$S^{(p)}(\tilde{\mathbf{x}}_r,\tilde{\mathbf{x}}_{s,n}) = g\left(\tilde{\mathbf{x}}_r^{(p)},\tilde{\mathbf{x}}_{s,n}^{(p)}\right) = Corr(\tilde{\mathbf{x}}_r^{(p)},\tilde{\mathbf{x}}_{s,n}^{(p)}) = \frac{\left[\tilde{\mathbf{x}}_r^{(p)}\right]\left[\tilde{\mathbf{x}}_{s,n}^{(p)}\right]^t}{\sqrt{\left[\tilde{\mathbf{x}}_r^{(p)}\right]\left[\tilde{\mathbf{x}}_r^{(p)}\right]^t}\sqrt{\left[\tilde{\mathbf{x}}_{s,n}^{(p)}\right]\left[\tilde{\mathbf{x}}_{s,n}^{(p)}\right]^t}} \tag{5}$$

meaning that $S^{(p)}(\tilde{\mathbf{x}}_r,\tilde{\mathbf{x}}_{s,n})$ is the similarity between the normalized and filtered derivatives (of order p) of the n'th ($n\in1\ldots N$) and m'th ($m\in1\ldots M$) UV-spectra in the HPLC data matrices, $\mathbf{X}_i$ and $\mathbf{X}_j$. If necessary, weights can be put on the spectrum in order to suppress or extract a specific range of wavelengths. Typically, if $M=N$ we also have $m=n$ due to the fact that the UV-spectra are compared sequentially.

Finally, each of the derivatives are combined through a convex linear combination

$$S_{ij,n} = \sum_{p=0}^{p} \omega_p S_{ij,n}^{(p)} \tag{6}$$

where ω_p is weight put on each derivative. In this study $\omega_p=1/P$ was chosen, which is equal to averaging the response from the derivatives.

2.6. Global similarity

The chromatograms are compared across the whole assembly of UV-spectra, one by one. This gives us a vector of similarities one for each spectrum

$$S = \left\{S_{ij,0},..,S_{ij,m},..,S_{ij,m}\right\} \tag{7}$$

To correct for unavoidable time drift due to different chemical and physical conditions between HPLC runs (a shift in retention time between two different runs) baseline correction and then aligning were applied. The baseline correction was performed on one wavelength at a time, by finding the minimum point in a 400 data point window for all possible window placements. Data points which were found as minimum more than twice were considered to be baseline points, and an estimated baseline for the current wavelength was created by linear interpolation between the detected baseline points. The resulting piecewise linear function was subtracted from the profile at the current wavelength, yielding a baseline corrected profile.

For aligning, several techniques can be applied (Pravdova et al., 2002) and here we have chosen the optimization method proposed by (Nielsen et al., 1998) due to the fact that this method has proven to give the globally optimal warp within the given set of a few parameters. The method aligns one chromatogram with another by dividing one chromatogram into small segments $[n_p;n_{p+1}]$ that may each be warped (e.g. stretched or compressed) to some degree. By using a global optimization method, a set of warps is found by the optimal set of new node positions that yields the best possible alignment to the other chromatogram. By aligning the profiles based on Eq. (5) we emphasize high spectral correlation in the alignment of the profiles. Within each segment we calculate the spectral correlations for all UV-spectra. Instead of warping the average chromatogram we apply the max–min strategy in order to find the best warp of the full chromatograms. This means that we maximize the minimal spectral correlation within each of the segments. This approach has shown to work very well, especially in those cases where many chromatographic peaks are present and we only align the profiles based on what is present along the spectral dimension.

The similarity between two HPLC matrices $\mathbf{X}_i$ and $\mathbf{X}_j$ is evaluated in two steps: 1) The similarity between

the UV-spectra where $\mathbf{X}_i$ has peaks, and 2) the similarity between the UV-spectra where $\mathbf{X}_j$ has peaks. The peaks were found as the mean absorbance value from each aligned spectrum, followed by a smothering using a simple mean filtering over a window of 9 UV-spectra. This was found to be both fast and sufficient to remove the small spikes that could be present; however, other filtering techniques could be applied (Eilers, 2003) if needed. Finally, we end up with two new vectors $\hat{S}_{ij}=\{\hat{S}_{ij,k}\}, k=1,\ldots, K$, and $\hat{S}_{ij}=\{\hat{S}_{ij,l}\}$, $l=1,\ldots, L$ where K and L now are the number of peaks detected in each of the matrices. The number of peaks can vary anywhere from 40 to 120 between the chromatograms. Experiments based on similarities between all UV-spectra, including those not regarded as peaks, did not give us good a result. Since $(\hat{S}_{ij}\neq\hat{S}_{ji})$ the scheme

$$\arg\min\left(\hat{S}_{ij},\hat{S}_{ji}\right) \tag{8}$$

at the 20% quantile is chosen to evaluate the overall similarity between $\mathbf{X}_i$ and $\mathbf{X}_j$.

The statistical calculations have been made using "R", a language and environment for statistical computing and graphics. R is available as Free Software, and can be downloaded from www.r-project.org. The software used for extracting data from the HPLC data files can be obtained together with a full documentation by contacting the corresponding author. Furthermore we have made the data available to the public from this site, since the authors are of the opinion that having a benchmark dataset is necessary in order to compare the performance of algorithms in the future.

3. Results

3.1. Method development

The results presented here are based on a method for unbiased/unsupervised chemotaxonomic classification of closely related fungi, using chemical

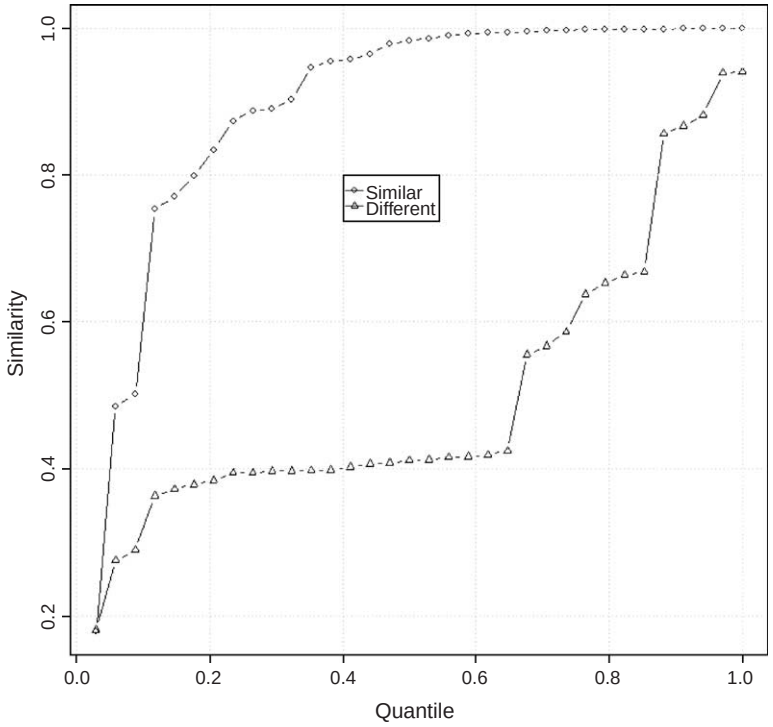

Fig. 2. Illustration of peak similarity between two similar isolates (-○-), *Penicillium atramentosum* (IBT 10565) and *P. atramentosum* (IBT 11801), and two different isolates (-△-), *P. atramentosum* (IBT 10565) and *P. camemberti* (IBT 21601).

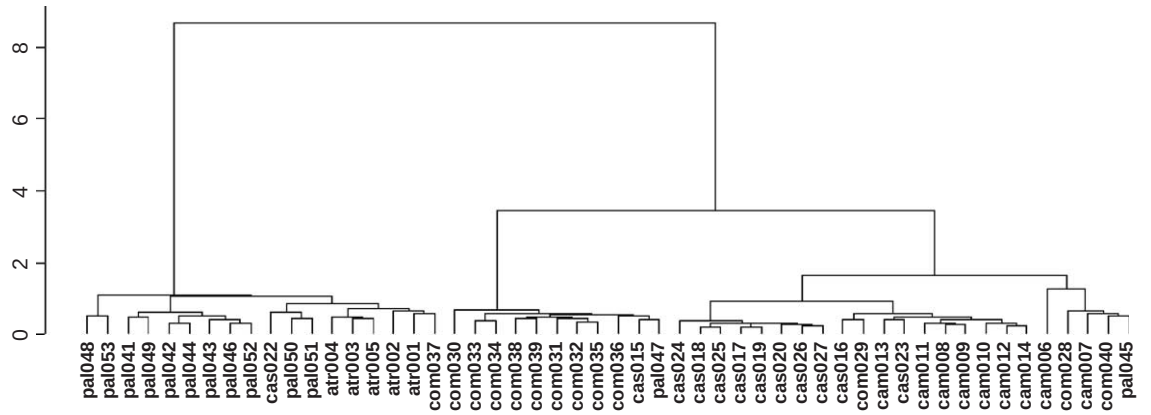

Fig. 3. Result of hierarchical clustering of 53 *Penicillium* isolates (CCC=0.78). The cluster analysis is done on the dissimilarity matrix using incremental sum of squares linkage (WARD). (atr: *P. atramentosum*; cam: *P. camemberti*; cas: *P. caseifulvum*; com: *P. commune*; pal: *P. palitans*).

analysis of secondary metabolite profiles. For all isolates, a similarity profile (see Eq. (7)) was calculated using the similarity described in Eq.(6). Each calculated profile represents "a best aligned match" between all HPLC data matrices along the retention time. The similarity profiles are finally evaluated through Eq. (8), giving an estimate of the overall similarity between the two profiles. In a similarity matrix, where profiles are more or less similar, the similarities between peaks in the profiles will have higher values, than where the profiles are different.

For two chromatographic data matrices a vector of locally aligned full spectral similarities was calculated along the retention time axis. For assessment of the chemotaxonomic grouping, a vector, evaluating the alikeness between two fungal extracts based upon each eluted compound and corresponding UV-spec-

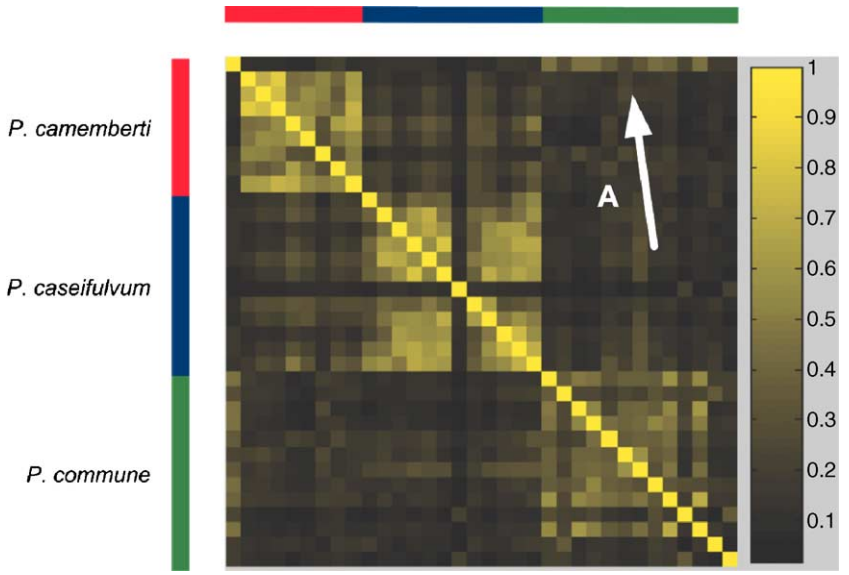

Fig. 4. The distance matrix of the three most closely related *Penicillium* species: *P. camemberti*, *P. commune* and *P. caseifulvum*. The matrix reveals one *Penicillium* isolate (IBT3505) that has been misidentified as *P. camemberti*.

trum, was first calculated and condensed into one similarity, depicting the overall degree of similarity between the profiles.

Fig. 2 shows the quantiles of the similarities between peaks in *P. atramentosum* (IBT10565) compared with the isolates *P. atramentosum* (IBT11801) and *P. camemberti* (IBT21601). The plot shows that there is a clear difference between similar and different species. In order to find the optimal quantile-value at which the "identical species" segregates from the "different species" we calculate the average profiles for all identical and different pairs, and find the location at which those two profiles differ the most. This value was found to be approximately at the 20% quantile. Here the similarity between all of the identical species was $\hat{S}_{ij}>0.7$ whereas $\hat{S}_{ij}<0.5$ for those different.

From Dataset 1 we found that the above described method performed well for both YES and CYA, even though there was a tendency to get a slightly better result on CYA than YES. Fig. 3 shows the hierarchical clustering of all of the *Penicillium* isolates used. Two main clusters are formed. One with *P. palitans* and *P. atramentosum*, and one with the species *P. commune*, *P. camemberti*, and *P. caseifulvum*. This main cluster is expected due to the fact that *P. palitans* and *P. atramentosum* are very chemically different compared to *P. commune*, *P. camemberti*, and *P. caseifulvum*. *P. camemberti* is domesticated from *P. commune* and used in the cheese industry. *P. commune* is regarded as a typical contaminant in dairies. All species are known to be difficult to segregate by traditional means.

The linkage function used in the hierarchical clustering was chosen based on having the highest cophenetic correlation coefficient (CCC) (Cormack, 1971).

The isolates that fall in the wrong clusters were investigated manually. During this process it was found that some of the extracts were low in concentration (i.e. not enough fungal material), which affected the results in such a way, that the noisy background starts to have an influence on the results. None of the extracts were overloaded. Fig. 4 shows the distance matrix for the three closely related species *P. camemberti*, *P. commune* and *P. caseifulvum*. The matrix reveals that there is one *P. camemberti* isolate (marked with an "A" in Fig. 4) that is almost certainly misclassified as a *P. camemberti*.

3.2. Method validation

The method was tested on a different set of fungal extracts. *Alternaria* belongs to a different genus than *Penicillium* and therefore produces different metabolites. Furthermore, the *Alternaria* isolates were grown on a different medium and the HPLC conditions were different compared to the *Penicillium* species, but

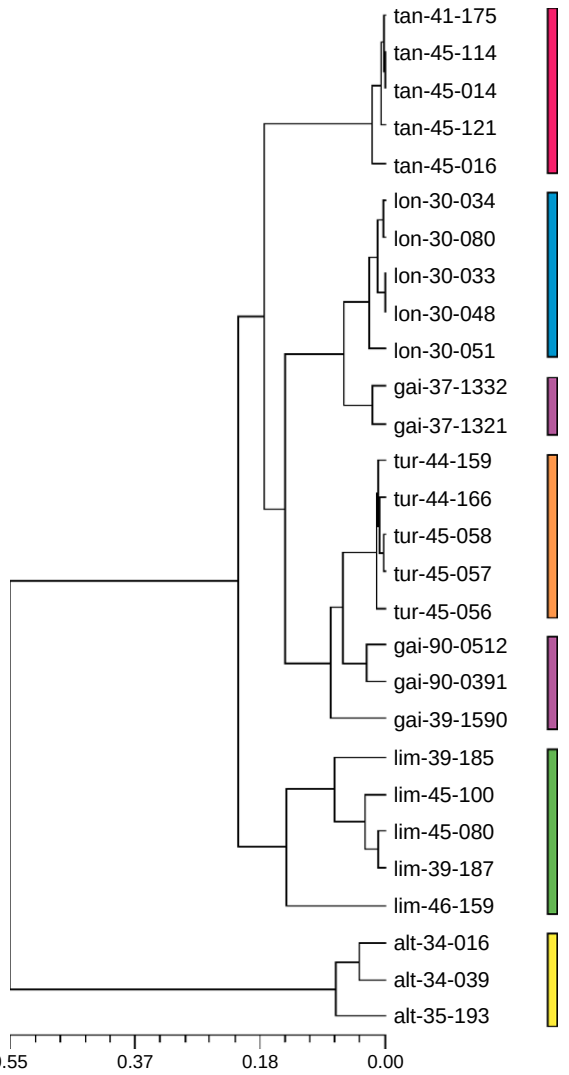

Fig. 5. Result of hierarchical clustering of 28 *Alternaria* isolates (CCC=0.83). The cluster analysis is done on the dissimilarity matrix using unweighted pair-group method, arithmetic average linkage (UPGMA). (alt: *A. alternaria*; gai: *A. gaisen*; lim: *A. limoniasperae*; lon: *A. longipes*; tan: *A. tangelonis*; tur: *A turkisafria*).

there were no changes in the parameters. Fig. 5 shows that the method segregated all species into species specific cluster, as predicted by the newest research, with the exception of two *A. gaisen* isolates. These *Alternaria* isolates have been regarded as the same species for many years based only on similar spore size. Manual exception of the two extracts that were located in the wrong cluster showed that the concentration again was too low and that background noise began to affect the result.

4. Discussion

As shown in this study, the above described method can be regarded as a powerful tool for classification and identification of complex fungal extracts. The method proposed here is both labour and time saving, although some of the algorithms are quite time consuming depending on the processing power available. Fortunately, research is currently going on leading to improved and even faster methods (Forshed et al., 2003; Pravdova et al., 2002). Also, computer power is constantly increasing leading to even shorter processing time.

One of the major advantages of applying the method is that the chemical diversity can be calculated by selecting only a few input parameters involved in the process: 1) Which part of the chromatograms and which spectral range to include, 2) perform a simple baseline correction, 3) align the chromatograms (to correct minor variations in retention times) by warping (Nielsen et al., 1998), 4) scale the chromatograms and finally 5) calculate the similarity. Therefore, the method removes the bias from comparisons and makes reproducibility possible between data files made at different periods. Most algorithms used for warping mainly rely on chromatographic traces, and problems may occur when several compounds are eluting within a short period of time. By using the full UV-spectral information in the aligning, the algorithm aligns peaks having the same UV-spectrum.

An improvement of the method will be to investigate and analyze the discriminating peaks. Another planned improvement is to include additional information about the extracts. This could be done by coupling a mass spectrometer (MS) to the output of the HPLC instrument, and hereby add mass spec. profiles to the UV-absorbance spectra. Of course the algorithm used to compare the mass spectra has to be different than for the UV-spectra due to the difference in the nature of data, but methods already exist for comparing such information (Hansen and Smedsgaard, 2004).

Acknowledgements

The authors thank Professor Jens Christian Frisvad (BioCentrum-DTU) for constructive discussions of the manuscript. Ellen Kirstine Lyhne and Hanne Jacobsen are acknowledged for cutting the plugs, extraction and analysis of the samples. The project was supported partly by the Danish Technical Research Council under the project: "Program for predictive biotechnology: Functional biodiversity in *Penicillium* and *Aspergillus*" (grant no. 9901295) and by the Danish Ministry of Food, Agriculture and Fisheries through the program "Food Quality with a focus on Food Safety".

References

Andersen, B., Nielsen, K.F., Thrane, U., Szaro, T., Taylor, J.W., Jarvis, B.B., 2003. Molecular and phenotypic descriptions of *Stachybotrys chlorohalonata* sp. nov. and two chemotypes of *Stachybotrys chartarum* found in water-damaged buildings. Mycologia 95, 1227–1238.

Cormack, R.M., 1971. A review of classification. Journal of the Royal Statistical Society. Series A. General 134, 321–367.

Counet, C., Ouwerx, C., Rosoux, D., Collin, S., 2004. Relationship between procyanidin and flavor contents of cocoa liquors from different origins. Journal of Agricultural and Food Chemistry 52, 6243–6249.

Egan, W.J., Morgan, S.L., bartick, E.G., Merril, R.A., Toylor, H.J., 2003. Forensic discrimination of photocopy and printer toners. II: discriminant analysis applied to infrared reflection–absorption spectroscopy. Analytical and Bioanalytical Chemistry 376, 1279–1285.

Eilers, P.H.C., 2003. A perfect smoother. Analytical Chemistry 75, 3631–3636.

Forshed, J., Schuppe-Koistinen, I., Jacobsson, S.P., 2003. Peak alignment of NMR signals by means of a genetic algorithm. Analytica Chimica Acta 487, 189–199.

Hansen, M.E., Smedsgaard, J., 2004. A new matching algorithm for accurate mass spectra. Journal of the American Society of Mass Spectroscopy 15, 1173–1180.

Nielsen, N.-P.V., Carstensen, J.M., Smedsgaard, J., 1998. Aligning of single and multiple wavelength chromatographic profiles for

chemometric data analysis using correlation optimised warping. Journal of Chromatography, A 805, 17–35.

Pitt, J.I., 1979. The genus *Penicillium* and its Teleomorphic States Eupencillium and Taleromyces. Academic Press, London.

Pravdova, V., Walczak, B., Massart, D.L., 2002. A comparison of two algorithms for warping of analytical signals. Analytica Chimica Acta 456, 77–92.

Samson, R.A., Hoekstra, E.S., Frisvad, J.C., Filtenborg, O., 2000. Introduction to food and airborne fungi, 6th ed. Centraalbureau voor Schimmelcultures, Utrecht.

Simmons, E.G., Roberts, R.G., 1993. *Alternaria* themes and variations. Mycotaxon 48 (73), 109–140.

Smedsgaard, J., 1997. Micro-scale extraction procedure for standardized screening of fungal metabolite production in cultures. Journal of Chromatography, A 760, 264–270.