Pyrolysis Mass Spectrometry as a Method for Inter-strain Discrimination of Candida albicans

By J. T. Magee, J. M. Hindmarch, B. I. Duerden and D. W. R. Mackenzie

1Department of Microbiology, The Children's Hospital, Sheffield S10 2TH, UK
2Department of Microbiology, Royal Hallamshire Hospital, Sheffield S10 2JF, UK
3Department of Medical Microbiology, The Medical School, University of Sheffield S10 2RX, UK
4Mycological Reference Laboratory, Central Public Health Laboratory, Colindale, London NW9 5HT, UK

(Received 15 January 1988; revised 28 June 1988)

A claim that Candida albicans strains NCPF 3153 and B311 were identical was investigated. Authentic strains were shown to be distinct ($P < 0.1\%$) by pyrolysis mass spectrometry (Py-MS). Of twelve strains, provided as clones of NCPF 3153, seven were authenticated, one yielded an equivocal result and four were distinct from both NCPF 3153 and B311. Of eight B311 clones, six were authenticated and two yielded equivocal results. Although five non-C. albicans yeast strains were identified as distinct from B311 and NCPF 3153, Torulopsis glabrata NCPF 3240 was identified as B311, and one clinical isolate of C. albicans as NCPF 3153. This could be explained by the specificity of the mathematical analysis for discrimination between the authentic strains.

INTRODUCTION

In pyrolysis, a sample of organic material is heated at a controlled temperature $> 300 \, ^{\circ}C$ in a non-oxidizing environment. Thermolytic cleavage of covalent bonds occurs, yielding a mixture of low molecular mass volatile compounds – the pyrolysate. The quantitative and qualitative composition of the pyrolysate reflects the pyrolysis conditions and composition of the sample, and these can be determined by gas–liquid chromatography (GC) or mass spectrometry. Pyrolysis of micro-organisms yields complex mixtures of products, giving fingerprint data which reflect the overall organic composition of the cell and the pyrolysis conditions. Identification, classification and typing of micro-organisms are possible with pyrolysis based characterization techniques (reviewed by Gutteridge & Norris, 1979; Drucker, 1981; Gutteridge et al., 1985; Shute et al., 1985). Pyrolysis mass spectrometry (Py-MS) offers greater resolution and speed than pyrolysis GC. However, commercial purpose-built Py-MS apparatus has only recently become available.

Two common reference strains of Candida albicans are NCPF 3153 and B311; it has been suggested that these represent different accession numbers for a single strain (Howard et al., 1986). However, records maintained at the Mycological Reference Laboratory (MRL) suggest otherwise. We compared authentic strains to determine if they represented one or two strains, and cultures returned from collections to determine if a system could be devised for future authentication.

Abbreviation: Py-MS, pyrolysis mass spectrometry.
METHODS

Strains and culture conditions. Nineteen investigators and culture collections in the UK, France and the USA were asked to provide cultures of Candida albicans NCPF 3153 and B311. Sixteen cultures were submitted. All isolates were streak purified and freeze dried. For routine maintenance, they were subcultured on glucose peptone agar. One investigator returned a culture as NCPF 3153, together with a recombinant clone of this organism, but later redesignated the latter as a clinical isolate of C. albicans. Investigation of one returned culture of NCPF 3153 showed four colonial variants. The returned strains therefore comprised eleven putatively identified as NCPF 3153, including the four colonial variants as distinct strains, eight putatively identified as B311 and one clinical isolate of C. albicans. A further nine cultures were obtained from the National Collection of Pathogenic Fungi (maintained at the Mycological Reference Laboratory, Colindale, London, UK) comprising C. albicans NCPF 3153 (two cultures), B311 (one culture), and one culture each of C. parapsilosis NCPF 3104, C. tropicalis NCPF 3111, C. krusei NCPF 3101, C. guilliermondii NCPF 3099, Torulopsis glabrata NCPF 3240 and Rhodotorula glutinis NCPF 3146.

For Py-MS, glucose peptone agar slopes of one authentic strain each of B311 and NCPF 3153 were labelled as such; slopes of the remaining 27 strains were given coded labels. These were subcultured on blood agar plates (Columbia Agar Base, Oxoid, with 5%, v/v, horse blood, Gibco) poured from a single batch of medium and incubated aerobically for 48 h at 37 °C.

Sample preparation and Py-MS conditions. Pyrolysis foils (Horizon Instruments) were washed in acetone, dried and heated to 180 °C for 3 h. Pyrolysis tubes (Horizon Instruments) were soaked in chromic acid, rinsed three times in distilled water, once in ethanol, and once in acetone, then dried and heated at 180 °C for 3 h. A single foil was inserted, with flamed forceps, into each tube so as to protrude about 6 mm from the mouth. The tubes were then heated at 180 °C for 3 h. Well separated colonies were picked from the blood agar cultures with a flamed straight wire, taking care to avoid fragments of agar. The colonies were smeared on to the protruding foil to give a uniform surface coating. The tubes were then heated at 80 °C for 5 min and stored in the dark over phosphorus pentoxide in a vacuum desiccator before analysis.

For analysis, the foils were tamped into the tubes with a flame cleaned stainless steel tool so as to lie 10 mm from the mouth. Viton O-ring collars (Horizon Instruments) were placed on the tubes, which were then processed automatically in batches of 20 on a Horizon Instruments PYMS 200X pyrolysis mass spectrometer. Curie point pyrolysis was at 530 °C for 4 s in vacuum. The expansion chamber, valve and collimating tube interfacing the pyrolysis tube with the mass spectrometer were heated at 130 °C. Products were ionized by collision with a crossing beam of low energy (25 eV) electrons, and the ions were separated in a quadrupole mass spectrometer which scanned the pyrolysate 160 times at 0.35 s intervals on initiation of pyrolysis. Integrated ion counts at unit mass intervals from 11 to 200 were recorded on floppy disk, together with pyrolysis sequence number and total ion count for each specimen. Ten replicate analyses were done on the two authentic strains, and three on each of the coded strains. More detailed descriptions of the apparatus, technique and principles involved were given by Aries et al. (1986). A typical mass spectrum is shown in Fig. 1.

Mathematical analysis. Visual analysis of spectra was not helpful; differences were subtle and obscured by variation in sample size and complex statistical effects. Sample size variations between spectra were corrected by iterative re-normalization (Huff et al., 1981). The sets of ten replicate spectra for the authentic strains were normalized by summing the ion counts for masses 50–140 – the normalization set – for each spectrum. Masses 11–49 were eliminated because such small fragments usually show poor ion count reproducibility and masses 141–200

Fig. 1. A typical pyrolysis mass spectrum of C. albicans. Quantitative ion counts for masses 11–200 (horizontal axis) are represented as vertical bars; the intensity at each mass may be read on the vertical axis.
were excluded because the low ion counts obtained were subject to large statistical fluctuations. The raw ion count for each mass was multiplied by $10^{6}$ and divided by the sum of ion counts of the normalization set for that spectrum. This corrected each spectrum to a constant total ion count of $10^{6}$ for the normalization set. The within-strain reproducibility of ion counts for each mass was then assessed by dividing the within-strain standard deviation of normalized ion counts by the strain mean normalized ion count to give the coefficient of variation (COV). The mean COV for the two authentic strains was calculated and masses showing a mean COV $> 0.15$ were eliminated from the normalization set. The spectra were re-normalized with the reduced normalization set, COV values calculated, and masses eliminated in three further cycles to yield a stable normalization set. This normalization set was used to correct all spectra for sample size variation before further analyses.

Data selection was based upon the level of univariate discrimination between the replicate data for the authentic strains. Masses for which ion counts showed a total:mean within-strain variance ratio $> 1.5$ were selected for the analysis set. This set comprised masses 51–60, 62, 67, 68, 72, 74, 81–84, 86, 94–100, 102 and 112.

Normalization and data selection were done by BASIC programs on a Prime 750 minicomputer.

Discriminant analysis (SPSSX User's Guide, 1983) was done on normalized ion counts of the analysis set, labelling the replicate spectrum data for each strain as a separate group. This analysis optimized Py-MS discrimination between strains, correcting for systematic intra-strain variance and covariance effects, and represented this optimized discrimination as coordinates on each of twenty-eight orthogonal axes. These numbered axes represented successively smaller proportions of the inter-strain discrimination achieved, the first seven representing 89%. The distance of individual spectra to hypothetical mean spectra (centroids) for the two authentic strains on the first seven axes of this discriminant space were then assessed. The discriminant analysis output included the coordinates for these centroids, and for individual spectra. If $C_{1}$–$C_{7}$ represent the coordinates of the centroid of an authentic strain on axes 1–7, and $S_{1}$–$S_{7}$ are the coordinates on axes 1–7 for the pyrogram under assessment, then the Euclidian distance ($D$, a measure of dissimilarity) between the pyrogram and the centroid can be calculated from the equation

$$D = \sqrt{(C_{1} - S_{1})^{2} + (C_{2} - S_{2})^{2} + (C_{3} - S_{3})^{2} + (C_{4} - S_{4})^{2} + (C_{5} - S_{5})^{2} + (C_{6} - S_{6})^{2} + (C_{7} - S_{7})^{2}}$$

The histograms of distances (Figs 3a and 4a) from individual authentic strain spectra to their centroid appeared to be approximately Gaussian, so the mean Euclidian distance plus two SD for authentic strain spectra was selected as the upper limit of $D$ for which a spectrum was regarded as indistinguishable from the authentic strain. Each coded strain was then identified on the basis of comparisons of the $D$ values for their spectra to these upper limits.

RESULTS

An ordination diagram of the spectral data on the first two canonical discriminant axes, which accounted for 59% of inter-strain discrimination, is shown in Fig. 2. At this low level of resolution, replicate spectra formed tight clusters. Three strains were obvious outliers; one ($R. glutinis$) was an extreme outlier, and two ($C. krusei$ and $C. tropicalis$) were less extreme. The authentic strains were clearly separated, and spectra of nine coded strains overlapped those of the authentic strains. However, spectra of a further fifteen coded strains formed a diffuse elongate group showing no obvious affinities with either authentic strain. The higher resolution attained in the mathematical analysis could not be represented on these two axes.

Histograms of Euclidian distances from individual spectra to the authentic strain centroids for the twenty-seven coded strains and two authentic strains are shown in Figs 3 and 4. These illustrate 89% of the total inter-strain discrimination achieved – that represented on the first seven canonical discriminant axes. The distances between individual authentic NCPF 3153 spectra and the centroid for NCPF 3153 showed an approximately normal distribution (Fig. 3a). The mean was 11.6 and the SD was 0.90, giving an upper limit for identification as NCPF 3153 of $D = 11.7$. The corresponding figures for B311 (Fig. 4a) were 9.7 and 1.0, giving an upper limit of 11.7. The difference between the means was about six SE indicating a significantly greater variation between spectra of the authentic strain of NCPF 3153 than between those of strain B311, i.e. a greater dispersion of replicate spectra of strain NCPF 3153, and therefore, probably a greater variation in organic composition for NCPF 3153. Despite this difference in dispersion, spectra of the authentic B311 and NCPF 3153 strains showed consistent statistically significant differences, allowing clear discrimination between these strains at $P < 0.01%$.

The histogram of distances from the B311 centroid for all coded strains (Fig. 4b) showed a clear minimum, corresponding roughly to the upper limit for identification selected. This was
Fig. 2. An ordination diagram of spectral data for the twenty-seven coded and two authentic strains. The axes represent the first two canonical discriminant functions, which together accounted for 59% of the discrimination between strains. ●, Spectra of non-B311, non-NCPF 3153 strains; ■, spectra of B311 strains; ▲, spectra of NCPF 3153 strains. Replicate spectra of the same strain are joined by lines, which form triangles in the case of coded strains. For the two authentic strains, only the outlying spectra are shown, and these are joined as polygons, which enclose the areas in which the remainder of the replicate spectra for these strains were found.

Evidence of a break in the data at the expected level of dissimilarity. However, the histogram of distances of all coded strains from the NCPF 3153 centroid (Fig. 3b) showed a secondary maximum coinciding with the upper limit for identification at $D \approx 13$ — a discouraging feature, indicating a lack of correspondence between the identification limit and similarity distribution minima. Alteration of the upper limit for identification for NCPF 3153 to 15 was considered, as this corresponded to a clear histogram minimum, but this was rejected as lacking a firm statistical basis.

The results of Py-MS identification in the blind trial, together with putative strain identity, are shown in Table 1. Of the twenty-seven coded strains, eight were identified as indistinguishable from NCPF 3153, seven as indistinguishable from B311, one was distinguishable from B311 but similar, eight were dissimilar from both authentic strains, and three were indistinguishable from both authentic strains (equivocal). The strain of *T. glabrata* was misidentified as B311, and the clinical isolate of *C. albicans* as NCPF 3153. The remaining five non-*C. albicans* strains were distinguished from NCPF 3153 and B311. Of twelve strains putatively derived from NCPF 3153, seven (including the coded authentic strain) were indistinguishable from the authentic strain, one yielded an equivocal identification, and four were distinct from
Fig. 3. The distances of individual spectra from the mean spectrum of the authentic NCPF 3153 strain are shown as stacked histograms. (a) Distances for spectra of the two authentic strains. (b) Distances for the spectra of the coded strains. □, Spectra of non-NCPF 3153, non-B311 strains; □, spectra of B311 strains; □, spectra of NCPF 3153 strains. The arrows indicate the upper limit of distance for identification as NCPF 3153.

Fig. 4. The distances of individual spectra from the mean spectrum of the authentic B311 strain are shown as stacked histograms. (a) Distances for spectra of the two authentic strains. (b) Distances for spectra of the coded strains. □, Spectra of non-NCPF 3153, non-B311 strains; □, spectra of B311 strains; □, spectra of NCPF 3153 strains. The arrows indicate the upper limit of distance for identification as B311.

Table 1. Results of Py-MS identification for the twenty-seven coded strains

<table>
<thead>
<tr>
<th>Py-MS result</th>
<th>Putative strain identity</th>
<th>NCPF 3153</th>
<th>B311</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCPF 3153</td>
<td></td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>B311</td>
<td></td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Equivocal</td>
<td></td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>12</td>
<td>8</td>
<td>7</td>
<td>27</td>
</tr>
</tbody>
</table>

both authentic strains. The latter four included a small colony variant from the returned culture that yielded four morphological types; the other three variants were identified in agreement with putative origin. Of eight strains putatively derived from B311, six were indistinguishable from the authentic strain and two yielded equivocal identifications.
A second study was undertaken, in which those strains that had been identified equivocally or in disagreement were re-pyrolysed in comparison with the authentic strains. The three strains which yielded equivocal results initially were authenticated; the four strains putatively derived from NCPF 3153 but distinct from the authentic strain in Py-MS were again distinct, and the clinical isolate remained indistinguishable from NCPF 3153. T. glabrata was distinct from B311 in this study, although at a high level of similarity.

DISCUSSION

The B311 and NCPF 3153 reference strains showed clear differences in Py-MS and most coded strains were authenticated. Indeed, Py-MS can discriminate between strains of the same species for a wide range of micro-organisms including Listeria monocytogenes serotypes (Eshuis et al., 1977), hospital strains of Klebsiella (Meuzelaar et al., 1982), Bacillus thuringiensis (Gutteridge et al., 1985), and Bacillus cereus, in which Shute et al. (1986) showed that Py-MS data could be used to distinguish strains associated with the production of emetic or diarrhoeal food poisoning symptoms. Py-MS studies of Haemophilus ducreyi, multi-resistant Staphylococcus aureus and capsule type B Haemophilus influenzae done at Sheffield (J.T.M. and J.M.H., unpublished data) also showed inter-strain discrimination. The possible applicability to a wide range of micro-organisms and speed of the technique make Py-MS a potential tool for investigation of hospital cross-infection and community outbreaks.

The deliberate inclusion of non-Candida albicans strains and the clinical isolate tested the capabilities of the analysis when challenged with extraneous organisms – an important consideration in typing. The mathematical approach selected only those masses (15-2% of the data) with low ion count variability within, and high discrimination between these strains. This highly specific approach carried the risk of discarding data pertinent to discrimination of other strains or species at an early stage, with resulting mis-identification. The results suggest that preliminary screening to eliminate extraneous organisms is required in Py-MS typing, as in other strain comparison techniques.

The high rate of production of non-germinating variants noted by Howard et al. (1986) and the recovery of four colony variants from a returned culture of NCPF 3153 indicated that this strain was prone to variation. The significantly greater dispersion of NCPF 3153 spectra indicated greater inherent variation in composition between colony samples. The secondary maximum in the histogram of distances from the NCPF 3153 centroid for the coded strains suggested that significant clonal variation had occurred, and this was commented upon before the putative identities were revealed. Finally, the level of agreement between putative and Py-MS identity of the coded strains was lower for NCPF 3153 than for B311. This study indicates that NCPF 3153 is more prone to clonal variation than B311. Considerable compositional drift has evidently occurred in several lines derived from this strain. This emphasizes the dangers of assuming that derived clones maintained for long periods in culture collections will continue to exhibit properties identical to the parental strain.

We thank the Trent Regional Health Authority Research Committee for their support of earlier pyrolysis GC work which allowed us to develop the practical and mathematical approaches used here, and Perkin-Elmer Ltd for the loan of the PYMS 200X.

REFERENCES


