RESEARCH ARTICLE

Identification of harmless and pathogenic algae of the genus *Prototheca* by MALDI-MS

Martin von Bergen[1], Angelika Eidner[2]*, Frank Schmidt[1], Jayaseelan Murugaiyan[1], Henry Wirth[1,3], Hans Binder[3], Thomas Maier[4] and Uwe Roesler[2,5]*

[1] Department of Proteomics, UFZ – Helmholtz-Centre for Environmental Research, Leipzig, Germany
[2] Institute of Animal Hygiene and Veterinary Public Health, University of Leipzig, Leipzig, Germany
[3] Interdisciplinary Centre for Bioinformatics, University of Leipzig, Leipzig, Germany
[4] Bruker Daltonics, Bruker Daltonik GmbH, Leipzig, Germany
[5] Institute of Animal Hygiene and Environmental Health, Free University Berlin, Berlin, Germany

Received: November 29, 2007
Revised: November 22, 2008
Accepted: January 15, 2009

The only plants infectious for mammals, green algae from the genus *Prototheca*, are often overseen or mistaken for yeast in clinical diagnosis. To improve this diagnostical gap, a method was developed for fast and reliable identification of *Prototheca*. A collection of all currently recognized *Prototheca* species, most represented by several strains, were submitted to a simple extraction by 70% formic acid and ACN; the extracts were analyzed by means of MALDI-MS. Most of the peaks were found in the range from 4 to 20 kDa and showed a high reproducibility, not in absolute intensities, but in their peak pattern. The selection of measured peaks is mostly due to the technique of ionization in MALDI-MS, because proteins in the range up to 200 kDa were detected using gel electrophoresis. Some of the proteins were identified by peptide mass fingerprinting and MS^2 analysis and turned out to be ribosomal proteins or other highly abundant proteins such as ubiquitin. For the preparation of a heatmap, the intensities of the peaks were plotted and a cluster analysis was performed. From the peak-lists, a principal component analysis was conducted and a dendrogram was built. This dendrogram, based on MALDI spectra, was in fairly good agreement with a dendrogram based on sequence information from 18S DNA. As a result, pathogenic and nonpathogenic species from the genus *Prototheca* can be identified, with possible consequences for clinical diagnostics by MALDI-typing.

Keywords:
MALDI-typing / *Prototheca*

1 Introduction

Colorless algae of the genus *Prototheca* within the *Chlorellaceae* family are the only known plants that cause infections in humans and animals. The taxonomic status of *Prototheca* has been evolving in recent decades and five species are currently assigned to this genus: *P. zopfii*, *P. wickerhamii*, *P. blaschkeae*, *P. stagnora* and *P. ulmea* [1–3]. Furthermore, *P. zopfii* is divided into two genotypes (1 and 2) with different antigen patterns and biochemical features [3, 4]

Numerous studies have reported a pathogenic potential for *P. wickerhamii* and *P. zopfii*. The cases of human protothecosis are predominantly caused by *P. wickerhamii* and occur as local (predominantly cutaneous) and systemic infections mainly in immune-compromised patients, *e.g.* patients infected with HIV or treated with glucocorticoids [5–8]. *P. blaschkeae* were isolated from some cases of onychomycosis [3]. Canine protothecosis is caused by *P. wickerhamii* and *P. zopfii*, and is characterized by similar clinical symptoms as in humans [9].

Correspondence: Dr. Martin von Bergen, Department of Proteomics, UFZ – Helmholtz-Centre for Environmental Research, Permoserstr. 15, D-04318 Leipzig, Germany
E-mail: Martin.vonbergen@ufz.de
Fax: +49-341-235-1786

*These authors contributed equally to this work.

World-wide, *P. zopfii* has been identified to induce a therapy-resistant inflammation of the mammary gland in dairy cows, which may lead to severe economic losses in an infected herd. Here, isolates assigned to ''variant'' or genotype 2 were reported to be the infectious agents [10–13]. On the other hand, *P. zopfii* genotype 1 is also often isolated in cattle barns, and has to be designated only as a milk contaminant. Because of its epidemiological impact and because of the broadly occurring resistance of the pathogenic *Prototheca* isolates against antimycotics, the correct identification of *P. zopfii, P. blaschkeae* and *P. wickerhamii* isolates is of considerable importance in clinical microbiological laboratories.

The diagnosis of *Prototheca* spec. is still based upon the time-consuming cultivation on Sabouraud-dextrose-agar medium, and on the additional investigation of lactophenol cotton blue stained cells by light microscopy [14–16]. However, because of the *Candida*-like appearance of the grown colonies and slow growth of most *Prototheca* strains, these methods are uncertain. Comparative investigations by means of Fourier-transformed infrared spectroscopy showed distinct differences between all *Prototheca* species, including *P. blaschkeae*, the former ''variant 3'' of *P. zopfii*. However, the discrimination of the two *P. zopfii* genotypes (the former variants 1 and 2 [17]) remains difficult. *P. zopfii* genotypes can currently be differentiated by sequence analysis of the 18S rDNA, or by diagnostic PCR or RFLP [3]. Although this method is the most accurate available up to now, this test is not capable of differentiating *Prototheca sp.* from yeast and the different strains of *Prototheca* from each other in a single run.

This diagnostical gap can be overcome by using MALDI-TOF MS spectra for identification. The possibility of identifying microorganisms by a combination of MALDI-TOF MS spectroscopy and advanced statistical analysis was first reported about a decade ago [18–20]. In principle, samples are obtained by simple extraction and then measured by MALDI-TOF MS (for review see [21]). The used range of 3–20 kDa excludes smaller metabolites. The obtained peaks, however, do not necessarily represent proteins. On the one hand, MALDI-TOF MS strongly favors proteins and peptides smaller than 20 kDa with some dependence upon the matrix used [22]. In reference to the term ''proteome,'' the collection of proteinagous molecules falling in this molecular range are referred to as the peptidome [23]. These molecules are well suited for both reverse phase chromatography and MALDI-TOF MS, which were widely used to generate peptidome-wide approaches for analytical and diagnostical purposes [24]. On the other hand, the population of smaller peptides can be dominated by irreproducible variations of breakdown products, so that for biomarker discovery more stable proteins are preferred [25].

For identification of microorganisms on the basis of extracted peptides, these have to be of intermediate or high abundance in order to obtain reliable results. Furthermore, they have to be expressed constitutively, because even slight variance in growing conditions can have gross effects on protein expression patterns in microorganisms [26, 27]. One group of peptides that fulfil all these criteria are the ribosomal proteins such as the 56 ribosomal proteins of *E.coli*, which ranges from 4364 to 29 727 Da [28] and are constitutively expressed with quite high abundance. The suitability of ribosomal proteins for identification has been reported recently [21, 29].

The identification of microorganisms requires compilation of a database with the reference spectra of known species as a first step. There are commercially available databases that currently comprise spectra of approx. 1500 different species [30]. In this approach, a number of spectra from one species are applied to the BioTyper^{TM} software where the discriminating peaks are extracted. There is also academic freeware such as MS-Screener [31] and EXPANDER [32] available, which can be used for the evaluation of the data and are suitable for the purpose of identifying microorganisms.

The combination of MALDI-measurement of whole-cell lysates or extracts with PCA and cluster algorithms has been used for the identification of a wide range of microorganisms, covering Gram-negative and Gram-positive bacteria, fungi and cyanobacteria [33–39]. Here a method is presented which aims at identifying green algae of the genus *Prototheca* that can occur as infectious pathogens in animals and humans. To support diagnostic applications we developed an experimental protocol and built a database, which allowed identification of isolates of *Prototheca* in environmental and medical samples.

2 Materials and methods

2.1 Chemicals

All chemicals and solvents used were of pro analysis quality and purchased from Sigma (Taufkirchen, Germany) or Merck (Darmstadt, Germany). High purity water was produced by an Ultra Clear UV plus system from SG GmbH (Barsbüttel, Germany).

The MALDI matrix solution was prepared by dissolving 6 mg CHCA in 50% ACN/2.5% TFA in H_2O.

2.2 Culturing of algae strains

A total of 19 strains, representing 7 algae species (5 *Prototheca* spec. and 2 *Chlorella* spec.) and 2 genotypes of *Prototheca zopfii* were used in this study (Table 1). All strains were either type, reference, or other well-characterized isolates that had previously been identified by sequence and biochemical analysis.

All strains were maintained on Sabouraud dextrose agar (Merck) plates, excepting *C. vulgaris*, which was

Table 1. Used *Prototheca* species, strains, and isolates

Taxa	Strains	
	Designation	Source/clinical signs
Prototheca zopfii Genotype 1	PZ I-1 [E, R, a)]	Pig manure, Germany
Prototheca zopfii Genotype 1	PZ I-2 [E, R, a)]	Pig manure, Germany
Prototheca zopfii Genotype 1	SAG 2063 [T, E, a) b)]	Pig manure, Germany
Prototheca zopfii Genotype 1	Hey1 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	Hey 5 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	Hey 7 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	Hey 9 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	Hey 10 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	Hey 12 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	Hey 13 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	Hey 15 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	POT 1 [C, a)]	Bovine mastitis, Germany
Prototheca zopfii Genotype 1	Gron 1 [E, a)]	Pig feces, Germany
Prototheca zopfii Genotype 1	SAG 263-7 [a) b) c) d)]	Environment (?)
Prototheca zopfii Genotype 1	SAG 263-10 [a) b) c) d)]	Sap from wounded *Robinia pseudoacacia*
Prototheca zopfii Genotype 2	SAG 2021 [T, C, a) b)]	Bovine mastitis, Germany
Prototheca zopfii Genotype 2	PZ II-2 [R, C, a)]	Bovine mastitis, Germany
Prototheca zopfii Genotype 2	PZ II-3 [R, C, a)]	Bovine mastitis, Germany
Prototheca zopfii Genotype 2	InsB [C, a)]	Human systemic infection (HIV assoc.), Austria
Prototheca zopfii Genotype 2	POT 2 [C, a)]	Bovine mastitis, Germany
Prototheca zopfii Genotype 2	POT 3 [C, a)]	Bovine mastitis, Germany
Prototheca zopfii Genotype 2	POT 4 [C, a)]	Bovine mastitis, Germany
Prototheca zopfii Genotype 2	DAW [C, a)]	Rectal swab from dog with chronic colitis, USA
Prototheca zopfii Genotype 2	Bras 1A [C, a)]	Bovine mastitis, Brazil
Prototheca zopfii Genotype 2	Bras 8 [C, a)]	Bovine mastitis, Brazil
Prototheca zopfii Genotype 2	Bras 14 [C, a)]	dog with hemorrhagic enterocolitis, Brazil
Prototheca zopfii Genotype 2	LAG 1 [C, a)]	Bovine mastitis, Belgium
Prototheca zopfii Genotype 2	SAG 263-3 [C, a) b) d)]	Human intestine
Prototheca zopfii Genotype 2	SAG 263-4 [C, a) b) d)]	Human intestine
Prototheca zopfii Genotype 2	SAG 263-1 [a) b) d)]	Sap from wounded tree
Prototheca blaschkeae	PZ III-1 [R, E, a)]	Cattle manure, Germany
Prototheca blaschkeae	PZ III-2 [R, E, a)]	Cattle manure, Germany
Prototheca blaschkeae	SAG 2064 [T, C, a) b)]	Human onychomycosis, Germany
Prototheca blaschkeae	PZ III-4 [E, a)]	Cattle manure, Germany
Prototheca blaschkeae	PZ III-5 [E, a)]	Cattle manure, Germany
Prototheca blaschkeae	PZ III-6 [E, a)]	Cattle manure, Germany
Prototheca blaschkeae	PZ III-7 [E, a)]	Cattle manure, Germany
Prototheca blaschkeae	PZ III-8 [E, a)]	Cattle manure, Germany
Prototheca blaschkeae	Au 66 [C, a)]	Bovine mastitis, Germany
Prototheca blaschkeae	Au 67 [C, a)]	Bovine mastitis, Germany
Prototheca blaschkeae	Au 70 [C, a)]	Bovine mastitis, Germany
Prototheca blaschkeae	LAG 10 [C, a)]	Bovine mastitis, Belgium
Prototheca wickerhamii	ATCC 16529 [T, E, c)]	Human sewage
Prototheca wickerhamii	CBS 157.74 [C, e)]	Human systemic infection
Prototheca wickerhamii	CBS 344.82 [C, e)]	Human dermatitis
Prototheca wickerhamii	RW 1 [C, a)]	Human dermatitis, Germany
Prototheca wickerhamii	RW 2 [E, a)]	Environment, Germany
Prototheca wickerhamii	RW 2a [E, a)]	Environment, Germany
Prototheca wickerhamii	RW 3 [E, a)]	Environment, Germany
Prototheca wickerhamii	ATCC 30395 [C, c)]	Human palmar lesions
Prototheca wickerhamii	BuW [C, a)]	Human gastroenteritis
Prototheca stagnora	ATCC 16528 [T, E, c)]	Sludge, Lebanon
Prototheca stagnora	UTEX 1442 [E, d)]	Digested sludge, USA
Prototheca ulmea	ATCC 50112 [T, E, c)]	Sap from wounded *Ulmus americana*, USA

Table 1. Continued

Taxa	Strains	
	Designation	Source/clinical signs
Chlorella protothecoides	ATCC 30407 [E, c)]	Sap from wounded *Populus alba*
Chlorella protothecoides	UTEX 249 [E, d)]	Environment
Chlorella protothecoides	SAG 211-10a [E, b)]	Sap from wounded *Ulmus* sp.
Chlorella saccharophila var. *ellipsoidea*	SAG 211-1a [T, E, b)]	Freshwater

T, Type strain; R, Reference strain; C, Clinical isolate; E, Environmental isolate.
a) Culture Collection of the Institute of Animal Hygiene and Veterinary Public Health of the University Leipzig, Leipzig, Germany.
b) Culture Collection of Algae at the University of Göttingen (SAG), Göttingen, Germany.
c) American Type Culture Collection (ATCC), Manassas, USA.
d) The Culture Collection of Algae at The University of Texas (UTEX), Austin, USA.
e) Centraalbureau voor Schimmelcultures (CBS), Baarn, Neederlands.

cultured on Bold's Basal Medium. The *Prototheca* strains were incubated at 25°C in the dark. The photosynthetic algae *A. protothecoides* and *C. vulgaris* were cultured at room temperature under light exposure.

2.3 Preparation of protein extracts

A standard protocol, in which colonies were directly applied to a MALDI-target [30], was modified by introducing a washing step. In brief, isolated colonies from agar plates were suspended in 300 μL water followed by the addition of 900 μL absolute ethanol. After centrifugation at 12 000 rpm for 2 min the supernatant was removed. The pellet was washed three times by cycles of resuspension and centrifugation in 1 mL water. The final cell pellet was dissolved in 50 μL of 70% formic acid followed by the addition of 50 μL ACN and thorough mixing. The suspension was centrifuged at 12 000 rpm in a bench-top centrifuge for 5 min at room temperature, then 0.5 μL of the clear supernatant was spotted in duplicates onto the MALDI target (MTP AnchorChip, Bruker Daltonik, Bremen, Germany). After air drying, each sample was overlaid with 1.5 μL of CHCA matrix solution and allowed to dry for several minutes before the MALDI-TOF MS measurement.

2.4 SDS-PAGE

For 1-D SDS-PAGE [40], 50 μg protein were precipitated with a final concentration of 10% trichloroacetic acid and washed twice with acetone [41]. The final pellets were redissolved in 10 μL sample buffer, heated for 5 min at 60°C and applied to the gel (4% acrylamide concentration in stacking and 12% acrylamide concentration in separating gel); the protein was visualized by silver staining [42]. For separation of peptides, 30 μg of protein extract was applied to a peptide gel (4% acrylamide in stacking, 10% acrylamide in spacer gel and 16.5% in separating gel) and stained by Coomassie Blue.

2.5 MALDI-MS for biotyping and protein identification by nano-LC-ESI-MS

A standard protocol, in which colonies were directly applied to a MALDI-target [30], was modified by introducing a washing step. In brief, isolated colonies from agar plates were suspended in 300 μL water followed by the addition of 900 μL absolute ethanol. After centrifugation at 12 000 rpm for 2 min the supernatant was removed. The pellet was washed three times by cycles of resuspension and centrifugation in 1 mL water. The final cell pellet was dissolved in 50 μL of 70% formic acid followed by the addition of 50 μL ACN and thorough mixing. The suspension was centrifuged at 12 000 rpm in a bench-top centrifuge for 5 min at room temperature, then 0.5 μL of the clear supernatant were mixed with DHB and spotted in duplicates onto a MALDI target (MTP ground steel, Bruker Daltonik).

Protein bands of interest were cut from polyacrylamide gels and digested overnight using trypsin (Sigma, Germany) as described elsewhere [42]. The cleaved peptides were eluted, concentrated by vacuum centrifugation and then separated by RP nano-LC (LC1100 series, Agilent Technologies, Paolo Alto, California; column: Zorbax 300SB-C18, 3.5 μm, $150 \times 0.075\,mm^2$; eluate: 0.1% formic acid, 0–60% ACN). The peptides were identified by on-line MS/MS (LC/MSD TRAP XCT mass spectrometer, Agilent Technologies) [43]. Subsequently, a database search was conducted using the MS/MS ion search (MASCOT, http://www.matrixscience.com) against all fungal entries of NCBInr (GenBank) with the following parameters: trpysin digestion, up to one missed cleavage, fixed modifications: carbamidomethyl (C), and with the following variable modifications: oxidation (M), peptide tol.: ±1.2 Da, MS/MS tol.: ±0.6 Da and peptide charge: +1, +2 and +3.

2.6 Data analysis

Peaks were detected from the raw mass spectra using the centriod algorithm from FlexAnalysis 2.4 with $S/N = 6$

(Bruker Daltonics) where only the highest 100 peaks were labeled to normalize the spectra. The generated peak lists were exported and processed using the MS-Screener program (Version 1.0.1) [31]. To aggregate the peak lists into matrices for subsequent generation of heatmaps the Expander [32] software was used; this software is based on cluster analysis using PAST [44].

Hierarchical agglomerative clustering was performed based upon mass spectra binned with an accuracy of 3.0 Da and preprocessing using quantile normalization. The distance matrix between the spectra was calculated using the average linkage metrics based on Pearson correlation coefficients. The hierarchical agglomerative algorithm successively merges the closest (most similar) objects (spectra or groups of spectra). The final number of clusters was visualized as a color-coded heatmap (see Supporting Information). The algorithm was implemented in C++ (program code is available on-line at http://izbifs.izbi.uni-leipzig.de/~wirth/Prototheca.zip).

For comparison between MALDI-MS based biotyping and genomic similarity measures, the 18S rDNA sequences of the algae strains investigated were analyzed as recently described [3]. In particular, we sequenced the 18S rDNA of the *Prototheca* strains studied, then performed multiple sequence alignment using CLUSTAL W [45] with the MegAlign™ module of the LASERGENE® software package (DNASTAR, Madison, WI, USA). The available 18S rDNA reference sequences (GenBank) of some of the investigated algae strains were included under the accession numbers indicated in the tree (Fig. 4A). Finally, the aligned sequences were analyzed by the neighbur-joining method [46] using the TREECON© software without distance estimation.

3 Results

3.1 Reproducible production and measurement of protein extracts from algae strains

Any identification of pathogenic organisms based on MALDI-spectra requires a simple and robust sample preparation and reliable spectra. For sample preparation we added one simple washing step to a well-established protocol [30]. To test the reliability we analyzed the sources of variance by conducting technical and biological replicates. The technical replicates showed only minor deviations, whereas the biological replicates exhibited larger variances especially in terms of peak intensities. To reduce the technical variances, each measurement was performed in four replicates as standard procedure. The peak intensities obtained (*e.g.* the peaks at 6590.32, 8012.78 and 8806.49 *m/z*) vary by about 25% between the replicates. These variances are only relevant if the peak intensity is of utmost importance for classification; they can be neglected if different species show completely different peaks. In the cross-species comparison (Fig. 1) the differences in the peak pattern as such in the spectra become obvious. The spectra obtained from *P. blaschkeae* (SAG 2064; Fig. 3), *P. zopfii* Genotype 1 (SAG 2063; Fig. 1C) and *P. zopfii* Genotype 2 (SAG 2021; Fig. 1D) differed slightly in intensity, but more importantly they showed clear differences in the peak patterns. Beside the peaks detected in MALDI-MS, which almost certainly represent proteins, a great number of proteins with higher molecular weights were also shown to be present in the extracts by 1-D SDS-PAGE (data not shown).

3.2 The extraction protocol yielded a wide range of proteins; identification of proteins in the range from 8 to 20 kDa revealed highly abundant ribosomal proteins

To validate the influence of the extraction method on the composition of proteins, samples were subjected to 1-D PAGE, suitable for separation of proteins from 10 to 200 kDa (Fig. 2). The gel was stained by silver because of the low protein content. The bands obtained range from below 10 to about 120 kDa, whereas the MALDI-spectra show a much narrower peak region (3–20 kDa) owing to the selectivity of the ionization process in MALDI-MS, which favors smaller peptides. Some of the sample extracts were applied to a peptide gel to assign the observed peaks (Fig. 4B). The bands of interest, with an apparent molecular weight of 10–20 kDa, were cut off and analyzed using mass spectrometry to identify the respective peptides

The eight most abundant proteins detected in the range of 8–20 kDa provide evidence for ribosomal proteins (L30, S15, S14, S11, S16, L21 and L12) and ubiquitin (Table 2). All assignments were performed on the basis of cross-species comparisons because of the lack of a sequenced *Prototheca* species. *Ostrecoccus tauri* seems to be the most closely related species with three hits in the MASCOT search, followed by *Chlamydomonas incerta*. The quality of this identification method is sufficient to claim unambiguous identification only for the first six proteins, due to the lack of sequence information about nearest phylogenetic relatives. For the ribosomal proteins L21 and L12, the measurement of only one peptide qualifies these identifications as likely, because the MASCOT search was performed against the complete NCBI database.

3.3 Differentially expressed proteins allow discrimination of *Prototheca* species

Based on the 18S rDNA phylogenetic tree, the quality of the protein-based phylogenetic trees was validated. 18S rDNA sequence data for all *Prototheca* and *Chlorella* strains investigated were available from GenBank or were achieved and presented in a recent study [3]. In the 18S rDNA-based phylogeny, the strains of the two genotypes of *P. zopfii* as well as of *P. blaschkeae* and *P. wickerhamii* forms consistent individual clades and the investigated strains show a high degree of homology within their genotype or species.

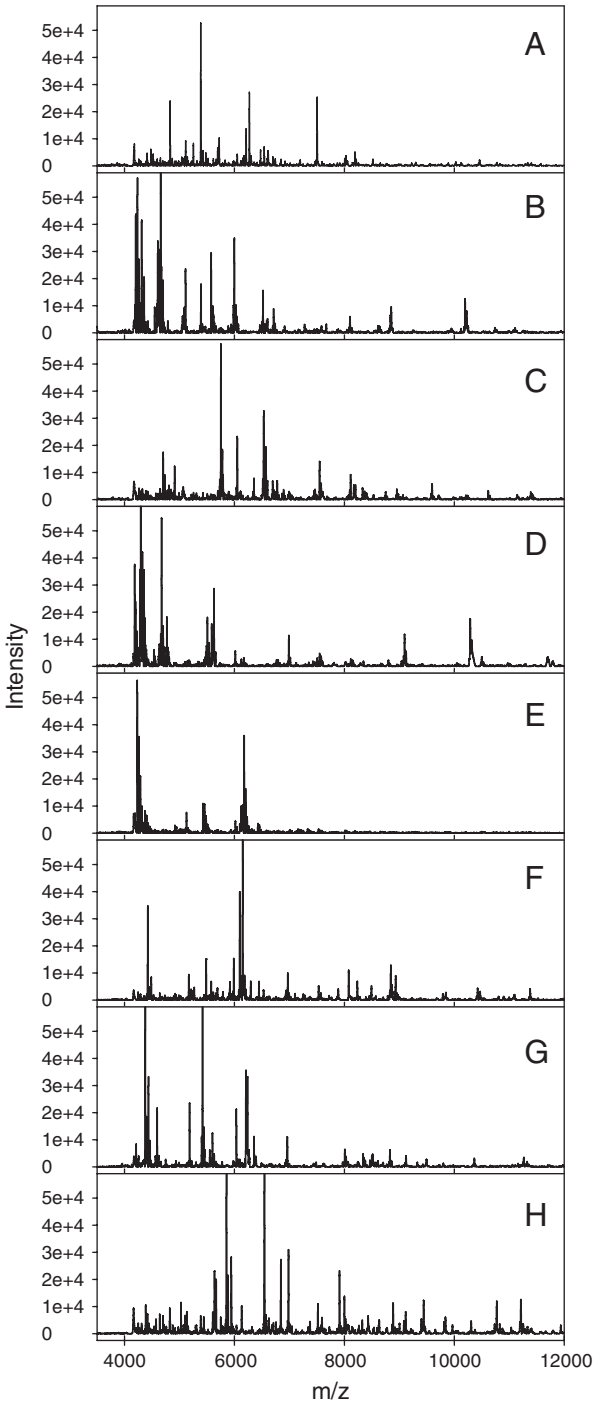

Figure 1. Comparison of spectra of different species. Representative MS-spectra in the range from 3.5 to 12 kDa from different strains as follows: (A) *Prototheca wickerhamii*; (B) *Prototheca stagnora*; (C) *Prototheca wickerhamii*; (D) *Chlorella protothecoides*; (E) *Prototheca zopfii* Genotype 2; (F) *Prototheca zopfii* Genotype 1; (G) *Prototheca blaschkeae* and (H) *Prototheca ulmea.*

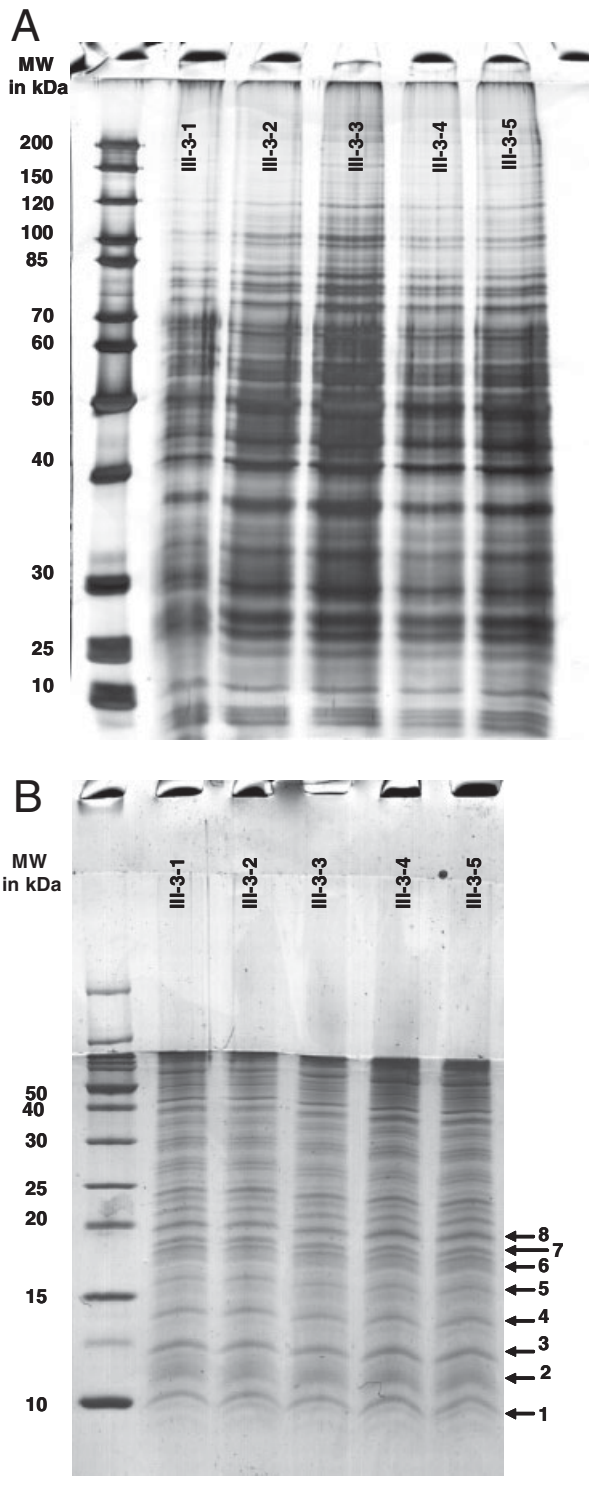

Figure 2. Separation of protein extracts by SDS-PAGE. Protein extracts from *Prototheca blaschkeae* used for MALDI-analysis were applied to 10% (A) and 16.5% (B) acrylamide concentration in the separation gel. The gels were stained with silver (A) and with Coomassie Blue (B). The bands indicated with arrows were cut and the proteins in them identified.

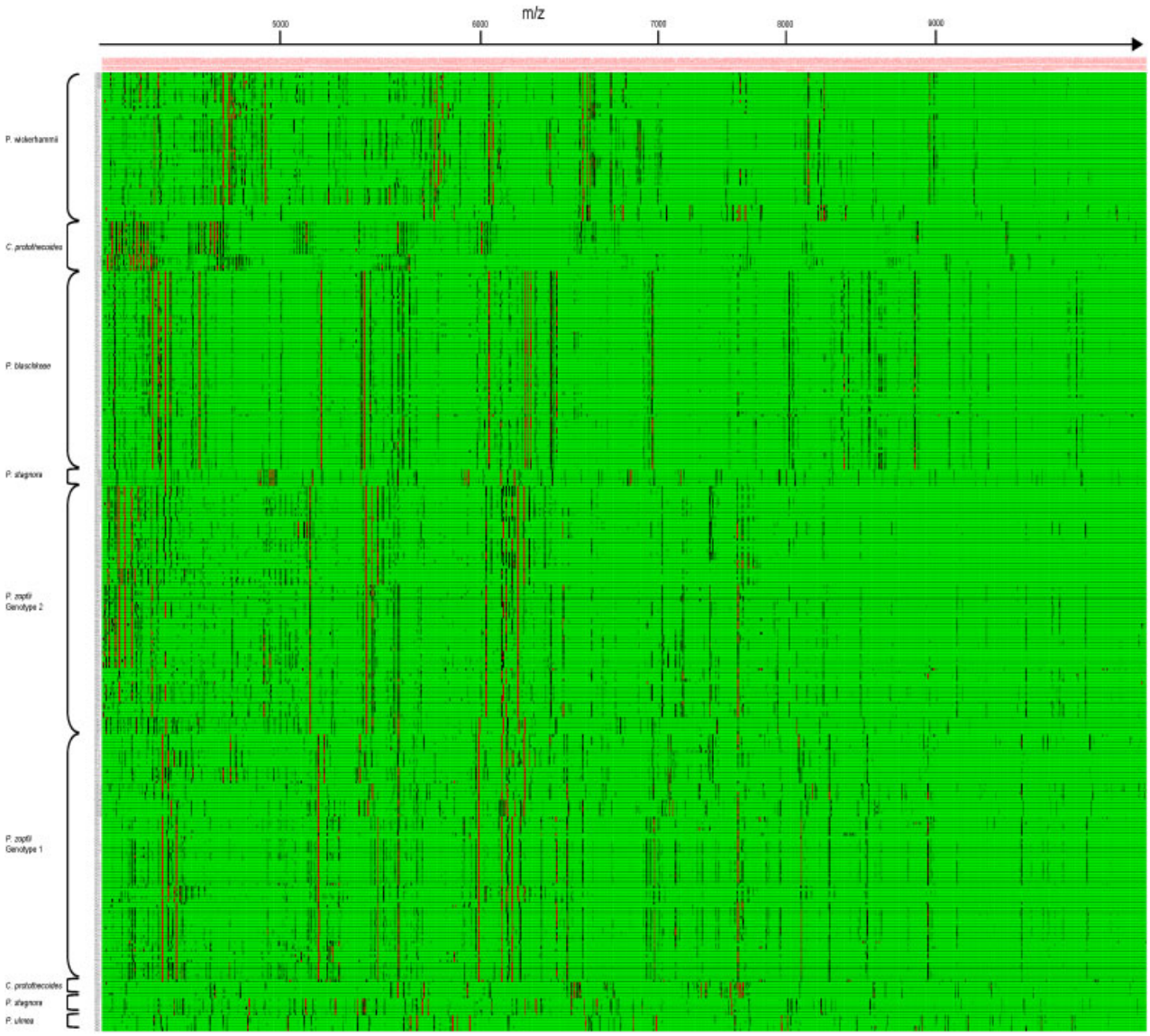

Figure 3. Heatmap of MS-spectra. The spectra are binned with an accuracy of 3.0 Da and the intensity is plotted in terms of colours, whereby red represents high and green low intensity.

For the preparation of the protein-based phylogenetic tree, first a heatmap of the presence and intensity of the peaks was generated (Fig. 3), based upon which the clustering by Pearson correlation was performed and compared with the DNA-based phylogenetic tree (Fig. 4).

This showed obvious features such as the high consistency of the clustering within the different species. The various spectra from one strain normally show a deviance of less than 20%. The spectra of the biological replicates varied from 0.06% (Goez6), typically 0.10% (Ulmea), up to 0.21% (CBS1), with an average of 0.11% over all samples.

Another general aspect is the perfect clustering of spectra belonging to a group of *Prototheca* species represented by more than three strains. This rule applies for the species *P. wickerhamii*, *P. blaschkeae*, *P. zopfii* Genotype 2 and *P. zopfii* Genotype 1. Within these species the deviance ranges from 0.70% (*P. zopfii* Genotype 2) *via* 0.82% (*P. zopfii* Genotype 1) to 0.86% (*P. wickerhamii and P. blaschkeae*), but these groups are clearly separated from each other and form coherent clusters. The picture is different for the two species *C. protothecoides* and *P. stagnora*. The spectra of these fall into two groups, from which one group of C. protothecoides shows a slightly higher similarity with the *P. zopfii* branch (node at 92%) and the other for the *P. wickerhamii* branch (node at 97% variance). On the basis of DNA sequences there is a higher similarity toward the *P. wickerhamii* branch of the phylogenetic tree.

A similar inconsistency was found for *P. stagnora* that falls into two clusters: one is an outlier-group in relation to most of the species, as also found on the basis of DNA

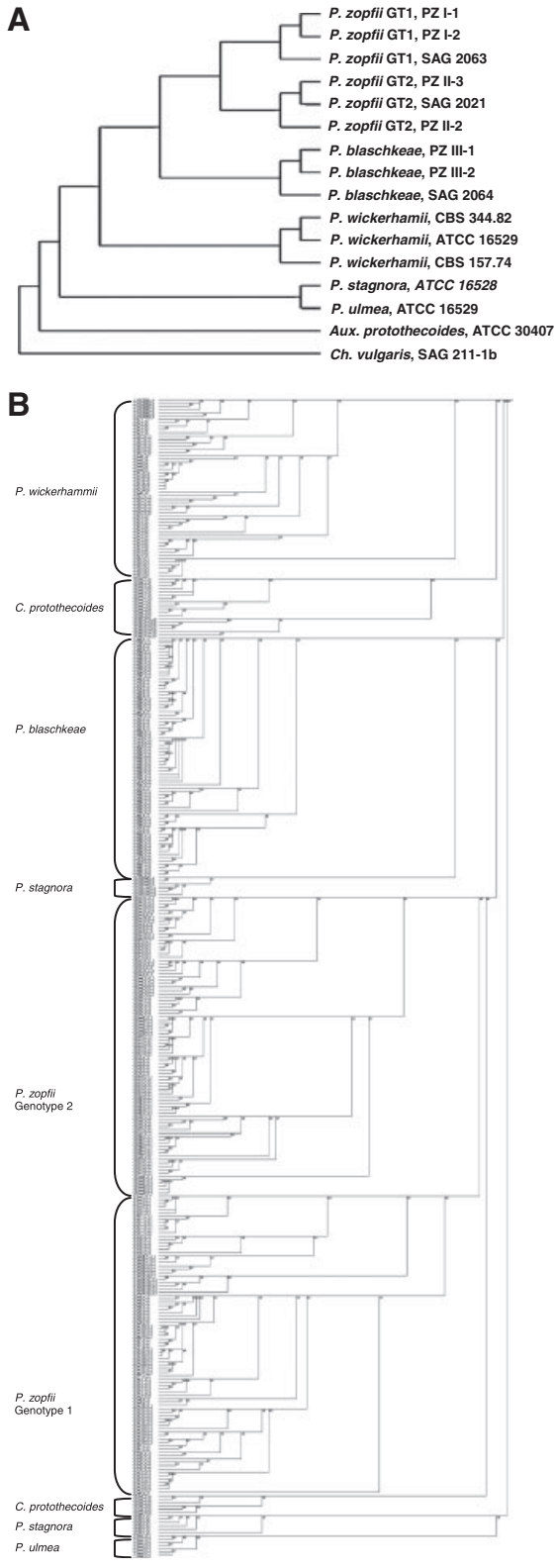

Figure 4. Comparison of a gene-based phylogenetic tree with a protein-based phylogenetic tree. The phylogenic tree based on 18SRNA (A) is shown in comparison to a phylogeny based on MS-spectra (B).

sequences, and the other showed 0.86 variance toward the *P. blaschkeae* cluster. The spectra of the first group stem from strain UTEX 1442 was isolated from digested sludge in the USA, whereas the other strain (ATC 16528) was isolated from sludge in Lebanon.

4 Discussion

4.1 Extension of MALDI-typing beyond bacteria

MALDI-typing was developed for the identification of bacteria (for review see [47]) but it seems likely that the method can be extended beyond the kingdom of eubacteria. Other studies have shown the suitability of this method for the phylum of cyanobacteria [35] or dermatophytes [48]. Here we have adapted the method to single cell algae from the kingdom of green plants. The main morphological difference to bacteria is obviously not their size, but the composition of the cell wall and cell membrane. Therefore, it is likely that initial failures to obtain meaningful spectra can be assigned to these features of algae cells. The standard sample preparation procedure is applicable for nearly all kinds of microorganisms similar to Gram-positive and Gram-negative bacteria and yeast [30]. Limitations were sometimes observed for organisms with very heavy, robust cell walls, such as certain filamentous fungi. Beside the problem of opening the cells and releasing the proteins of interest, cell surface modifications, such as lipids or carbo-hydrates, can also interfere with the MALDI measurement. In most cases, additional washing steps (either with 70% ethanol or with pure water) before the cell disruption are sufficient to remove troublesome substances. For very robust organisms, an additional boiling step before adding ethanol can improve MALDI measurements significantly. The results presented here open the path for using MALDI-typing for the identification of higher organisms such as plants and animals.

4.2 Comparison between classifications based on nucleotides and proteins

The "gold standard" in the identification of microorganisms has shifted in the past few decades from the morphological and physiological analysis to a nucleotide-based one since more and more sequence information and amplification techniques became available. Recently, the so-called barcoding approach has started to generate a more complete picture of evolutionary relationships between genera and species. Therefore, a set of genes was selected that has been sequenced in a great number of species from all taxa [49]. In this approach the 18S-rDNA genes are of utmost importance, although also the sequences from other proteins such as cytochrome oxidase were also used. However, sequencing is not a screening technique, so most often specific primers are used for PCR or RFLP analysis. These techniques have

Table 2. Identified proteins

Spot No.	Swiss-Prot Acc. No.	Entry name	Protein	Score	M_r	p*I*	Seq. cov. %	Pept. mat. No.
1	P42739	UBIQ_ACECL	Ubiquitin (*Acetabularia cliftonii*, green algae)	77	8536	6.56	36%	2
2	Q7XYA1	Q7XYA1_GRIJA	Ribosome protein L30 (*Griffithsia japonica*, red algae)	111	12410	9.75	10%	3
3	Q1WLZ3	Q1WLZ3_CHLIN	40S ribosomal protein S15 (*Chlamydomonas incerta*, green algae)	64	14936	10.05	13%	2
4	P46295	RS14_CHLRE	40S ribosomal protein S14 (*Chlamydomonas incerta*, green algae)	122	16349	10.31	15%	2
5	P42756	RS11_DUNTE	40S ribosomal protein S11 (*Dunaliella tertiolecta*, green algae)	85	17912	10.43	9%	4
6	Q018M7	Q018M7_OSTTA	40S ribosomal protein S16 (ISS) (*Ostreococcus tauri*, green algae)	109	18066	10.43	10%	2
7	Q00U75	Q00U75_OSTTA	60s ribosomal protein L21 (ISS) (*Ostreococcus tauri*, green algae)	61	19441	10.79	4%	1
8	Q01BX3	Q01BX3_OSTTA	putative 60S ribosomal protein L12 (ISS) (*Ostreococcus tauri*, green algae)	75	24761	11.11	6%	1

The proteins identified from the gel in Fig. 2 are summarized by listing the Swiss-Prot accession number, the uniprot accession number, the protein identification, and the MASCOT-MOWSE score as well as the molecular weight, p*I*, sequence coverage and number of peptides used for identifications. Except for the samples 7 and 8, for all other proteins at least two peptides were found; also for samples 7 and 8 significant MOWSE scores were achieved.

the potential for application as screening procedures, but reduce the amount of information from the sequence. Resources are limited, and although the technique of in-depth sequencing has been extensively developed, it is still too costly for such application, leaving a gap and a need for fast and reliable techniques for species identification. The technique of MALDI-typing has yielded valuable results for bacteria and the comparison to the ''gold standard'' of RNA-gene phylogeny showed that this holds also true for the identification of *Prototheca* species.

In principle, the informational content of a nucleotide sequence is threefold higher than that of the corresponding protein and, in addition, the protein can exist in multiple modifications not coded genomically. The development of mass spectrometers has increased the resolution space from about 50 peaks in the range from 400 to 2000 Da to that measured in this study by nearly two orders of magnitude by using nano-LC coupled instruments. Consequently, shotgun approaches are already capable of detecting more than 3000 proteins in a single run [50] and will probably also come into common use for species identification. Recent studies have shown the capacity of shotgun mass mapping to enable classification down to the strain level in the case of *Lactobacillus* [51]. This wealth of informational content will improve the resolution and reliability of MS-based identification. Beside the nontargeted approach of using the complete proteome species, identification can also make use of focussing on specific classes of proteins, *e.g.* membrane proteins [52] that are of special importance in pathogenic bacteria. The problem of condition-dependent protein expression is circumvented by taking most into account constitutively expressed proteins, and in addition by preparing a data set that contains master spectra derived from several reference spectra.

4.3 Differentiation and identification of pathogenic and nonpathogenic algae species

Based on the 18S rDNA phylogenetic tree, the quality of the protein-based phylogenetic approach based upon MALDI-MS was proven. The cladogram based upon MALDI-MS results resembled mostly the one based on 18S rDNA analysis. The inconsistencies, *e.g.* spectra of *P. stagnora* and *C. protothecoides* each falling into two different clusters, might be explained by the different heritage of the strains. In the case of *P. stagnora* one was isolated from digested sludge in the USA, the other from sludge in Lebanon. The different sources might have caused a different protein expression pattern (''Biotypes'') that remained even under changed cultivation conditions. Another possible reason is the missing 18S rDNA sequence data of the two outlier-group strains of *P. stagnora* (strain UTEX 1442) and *C. protothecoides* (strain SAG 211-10a). Therefore, it seems to be possible that these two strains are currently not assigned to their correct taxon.

However, apart from minor deviances from the DNA-sequence-based phylogenetic tree, the performed MALDI-MS analysis of intact proteins has been proven to allow a fast and accurate identification of the currently described pathogenic *Prototheca* species and genotypes (*e.g.* *P. wickerhamii*, *P. blaschkeae* and *P. zopfii* genotype 2) as well as of some closely related green algae species.

4.4 Conclusions

The described MALDI-MS analysis combined with unsupervised clustering provides a novel valuable tool for

detecting *Prototheca* species. In comparison to the classical microbiological tests this method is faster and offers a higher reliability, if performed by trained personnel. Moreover, MALDI-MS analysis involves lower costs and effort than the currently used genomic approaches (sequencing, PCR, RFLP) and might thereby support the clinical diagnosis of *Prototheca* infections.

We thank Michaela Risch for excellent technical assistance.

The authors have declared no conflict of interest.

5 References

[1] Arnold, P., Ahearn, D. G., The systematic of the genus *Prototheca* with a description of a new species *P. filamenta. Mycologica* 1972, *64*, 265–275.

[2] Pore, R. S., Prototheca taxonomy. *Mycopathologia* 1985, *90*, 129–139.

[3] Roesler, U., Moller, A., Hensel, A., Baumann, D., Truyen, U., Diversity within the current algal species *Prototheca zopfii*: a proposal for two *Prototheca zopfii* genotypes and description of a novel species, *Prototheca blaschkeae* sp. nov. *Int. J. Syst. Evol. Microbiol.* 2006, *56*, 1419–1425.

[4] Roesler, U., Scholz, H., Hensel, A., Emended phenotypic characterization of *Prototheca zopfii*: a proposal for three biotypes and standards for their identification. *Int. J. Syst. Evol. Microbiol.* 2003, *53*, 1195–1199.

[5] Matsuda, T., Matsumoto, T., Protothecosis: a report of two cases in Japan and a review of the literature. *Eur. J. Epidemiol.* 1992, *8*, 397–406.

[6] Bianchi, M., Robles, A. M., Vitale, R., Helou, S. *et al.*, The usefulness of blood culture in diagnosing HIV-related systemic mycoses: evaluation of a manual lysis centrifugation method. *Med. Mycol.* 2000, *38*, 77–80.

[7] Lass-Florl, C., Fille, M., Gunsilius, E., Gastl, G., Nachbaur, D., Disseminated infection with *Prototheca zopfii* after unrelated stem cell transplantation for leukemia. *J. Clin. Microbiol.* 2004, *42*, 4907–4908.

[8] Lass-Florl, C., Mayr, A., Human protothecosis. *Clin. Microbiol. Rev.* 2007, *20*, 230–242.

[9] Stenner, V. J., Mackay, B., King, T., Barrs, V. R. *et al.*, Protothecosis in 17 Australian dogs and a review of the canine literature. *Med. Mycol.* 2007, *45*, 249–266.

[10] Costa, E. O., Ribeiro, A. R., Watanabe, E. T., Melville, P. A., Infectious bovine mastitis caused by environmental organisms. *Zentralbl. Veterinarmed. B* 1998, *45*, 65–71.

[11] Jensen, H. E., Aalbaek, B., Bloch, B., Huda, A., Bovine mammary protothecosis due to *Prototheca zopfii*. *Med. Mycol.* 1998, *36*, 89–95.

[12] Janosi, S., Ratz, F., Szigeti, G., Kulcsar, M. *et al.*, Review of the microbiological, pathological, and clinical aspects of bovine mastitis caused by the alga *Prototheca zopfii*. *Vet. Q.* 2001, *23*, 58–61.

[13] Moller, A., Truyen, U., Roesler, U., *Prototheca zopfii* genotype 2: the causative agent of bovine protothecal mastitis? *Vet. Microbiol.* 2007, *120*, 370–374.

[14] Pore, R. S., Shahan, T. A., Pore, M. D., Blauwiekel, R., Occurrence of *Prototheca zopfii*, a mastitis pathogen, in milk. *Vet. Microbiol.* 1987, *15*, 315–323.

[15] Costa, E. O., Carciofi, A. C., Melville, P. A., Prada, M. S., Schalch, U., *Prototheca* sp. outbreak of bovine mastitis. *Zentralbl. Veterinarmed. B* 1996, *43*, 321–324.

[16] Baumgärnter, B., Vorkommen und Bekämpfung der Protothekenmastitis des Rindes im Einzugsgebiet des Staatlichen Veterinär-und Lebensmitteluntersuchungsamtes Potsdam. *Prak. Tierarzt.* 1972, *78*, 406–414.

[17] Schmalreck, A. F., Trankle, P., Vanca, E., Blaschke-Hellmessen, R., Differentiation and characterization of yeasts pathogenic for humans (*Candida albicans*, *Exophiala dermatitidis*) and algae pathogenic for animals (*Prototheca* spp.) using Fourier transform infrared spectroscopy (FTIR) in comparison with conventional methods. *Mycoses* 1998, *41*, 71–77.

[18] Claydon, M. A., Davey, S. N., Edwards-Jones, V., Gordon, D. B., The rapid identification of intact microorganisms using mass spectrometry. *Nat. Biotechnol.* 1996, *14*, 1584–1586.

[19] Holland, R. D., Wilkes, J. G., Rafii, F., Sutherland, J. B. *et al.*, Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 1996, *10*, 1227–1232.

[20] Krishnamurthy, T., Ross, P. L., Rajamani, U., Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 1996, *10*, 883–888.

[21] Fenselau, C., Demirev, P. A., Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.* 2001, *20*, 157–171.

[22] Hortin, G. L., The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin. Chem.* 2006, *52*, 1223–1237.

[23] Schulz-Knappe, P., Schrader, M., Zucht, H. D., The peptidomics concept. *Comb. Chem. High Throughput Screen* 2005, *8*, 697–704.

[24] Hu, L., Ye, M., Jiang, X., Feng, S., Zou, H., Advances in hyphenated analytical techniques for shotgun proteome and peptidome analysis—a review. *Anal. Chim. Acta* 2007, *598*, 193–204.

[25] Hortin, G. L., Jortani, S. A., Ritchie, J. C., Jr., Valdes, R., Jr., Chan, D. W., Proteomics: a new diagnostic frontier. *Clin. Chem.* 2006, *52*, 1218–1222.

[26] Vargha, M., Takats, Z., Konopka, A., Nakatsu, C. H., Optimization of MALDI-TOF MS for strain level differentiation of Arthrobacter isolates. *J. Microbiol. Methods* 2006, *66*, 399–409.

[27] Williams, M. D., Ouyang, T. X., Flickinger, M. C., Starvation-induced expression of SspA and SspB: the effects of a null mutation in sspA on *Escherichia coli* protein synthesis and survival during growth and prolonged starvation. *Mol. Microbiol.* 1994, *11*, 1029–1043.

[28] Suh, M. J., Limbach, P. A., Investigation of methods suitable for the matrix-assisted laser desorption/ionization mass spectrometric analysis of proteins from ribonucleoprotein complexes. *Eur. J. Mass Spectrom.* 2004, *10*, 89–99.

[29] Pineda, F. J., Antoine, M. D., Demirev, P. A., Feldman, A. B. *et al.*, Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. *Anal. Chem.* 2003, *75*, 3817–3822.

[30] Maier, T., Kostrzewa, M., Fast and reliable MALDI-TOF MS-based microorganism identification. *Chem. Today* 2007, *25*, 68–71.

[31] Schmidt, F., Schmid, M., Jungblut, P. R., Mattow, J. *et al.*, Iterative data analysis is the key for exhaustive analysis of peptide mass fingerprints from proteins separated by two-dimensional electrophoresis. *J. Am. Soc. Mass Spectrom.* 2003, *14*, 943–956.

[32] Sharan, R., Maron-Katz, A., Shamir, R., CLICK and EXPANDER: a system for clustering and visualizing gene expression data. *Bioinformatics* 2003, *19*, 1787–1799.

[33] Dickinson, D. N., La Duc, M. T., Haskins, W. E., Gornushkin, I. *et al.*, Species differentiation of a diverse suite of Bacillus spores by mass spectrometry-based protein profiling. *Appl. Environ. Microbiol.* 2004, *70*, 475–482.

[34] Dieckmann, R., Graeber, I., Kaesler, I., Szewzyk, U., von Dohren, H., Rapid screening and dereplication of bacterial isolates from marine sponges of the sula ridge by intact-cell-MALDI-TOF mass spectrometry (ICM-MS). *Appl. Microbiol. Biotechnol.* 2005, *67*, 539–548.

[35] Fastner, J., Erhard, M., von Dohren, H., Determination of oligopeptide diversity within a natural population of *Microcystis* spp. (cyanobacteria) by typing single colonies by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* 2001, *67*, 5069–5076.

[36] Ishida, Y., Nakanishi, O., Hirao, S., Tsuge, S. *et al.*, Direct analysis of lipids in single zooplankter individuals by matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* 2003, *75*, 4514–4518.

[37] Li, T. Y., Liu, B. H., Chen, Y. C., Characterization of Aspergillus spores by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2000, *14*, 2393–2400.

[38] van Baar, B. L., Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. *FEMS Microbiol. Rev.* 2000, *24*, 193–219.

[39] Pignone, M., Greth, K. M., Cooper, J., Emerson, D., Tang, J., Identification of mycobacteria by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. *J. Clin. Microbiol.* 2006, *44*, 1963–1970.

[40] Laemmli, U., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, *15*, 680–685.

[41] Santos, P. M., Roma, V., Benndorf, D., von Bergen, M. *et al.*, Mechanistic insights into the global response to phenol in the phenol-biodegrading strain *Pseudomonas* sp. M1 revealed by quantitative proteomics. *OMICS* 2007, *11*, 233–251.

[42] Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F. *et al.*, Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* 1996, *93*, 14440–14445.

[43] Kellner, H., Jehmlich, N., Benndorf, D., Hoffmann, R. *et al.*, Detection, quantification and identification of fungal extracellular laccases using polyclonal antibody and mass spectrometry. *Enzyme Microb. Technol.* 2007, *41*, 694–701.

[44] Hammer, Ø., Harper, D. A. T., Ryan, P. D., PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 2001, *4*, 9.

[45] Thompson, J. D., Higgins, D. G., Gibson, T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994, *11*, 4673–4680.

[46] Saitou, N., Nei, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 1987, *4*, 406–425.

[47] Fenselau, C., Demirev, P. A., Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.* 2001, *20*, 157–171.

[48] Erhard, M., Hipler, U. C., Burmester, A., Brakhage, A. A., Wostemeyer, J., Identification of dermatophyte species causing onychomycosis and tinea pedis by MALDI-TOF mass spectrometry. *Exp. Dermatol.* 2008, *17*, 356–361.

[49] Hajibabaei, M., Singer, G. A., Hebert, P. D., Hickey, D. A., DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* 2007, *23*, 167–172.

[50] Haas, W., Faherty, B. K., Gerber, S. A., Elias, J. E. *et al.*, Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell Proteomics* 2006, *5*, 1326–1337.

[51] Schmidt, F., Fiege, T., Hustoft, H. K., Kneist, S., Thiede, B., Shotgun mass mapping of Lactobacillus species and subspecies from caries related isolates by MALDI-MS. *Proteomics* 2009, *9*, 1994–2003.

[52] Schaller, A., Troller, R., Molina, D., Gallati, S. *et al.*, Rapid typing of *Moraxella catarrhalis* subpopulations based on outer membrane proteins using mass spectrometry. *Proteomics* 2006, *6*, 172–180.

