Appl Microbiol Biotechnol (2014) 98:5645–5654
DOI 10.1007/s00253-014-5686-7

APPLIED MICROBIAL AND CELL PHYSIOLOGY

Evaluation of MALDI-TOF mass spectrometry for identification of environmental yeasts and development of supplementary database

Bruna Carla Agustini · Luciano Paulino Silva · Carlos Bloch Jr. · Tania M. B. Bonfim · Gildo Almeida da Silva

Received: 3 January 2014 / Revised: 8 March 2014 / Accepted: 11 March 2014 / Published online: 1 April 2014

Abstract Yeast identification using traditional methods which employ morphological, physiological, and biochemical characteristics can be considered a hard task as it requires experienced microbiologists and a rigorous control in culture conditions that could implicate in different outcomes. Considering clinical or industrial applications, the fast and accurate identification of microorganisms is a crescent demand. Hence, molecular biology approaches has been extensively used and, more recently, protein profiling using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has proved to be an even more efficient tool for taxonomic purposes. Nonetheless, concerning to mass spectrometry, data available for the differentiation of yeast species for industrial purpose is limited and reference databases commercially available comprise almost exclusively clinical microorganisms. In this context, studies focusing on environmental isolates are required to extend the existing databases. The development of a supplementary database and the assessment of a commercial database for taxonomic identifications of environmental yeast are the aims of this study. We challenge MALDI-TOF MS to create protein profiles for 845 yeast strains isolated from grape must and 67.7 % of the strains were successfully identified according to previously available manufacturer database. The remaining 32.3 % strains were not identified due to the absence of a reference spectrum. After matching the correct taxon for these strains by using molecular biology approaches, the spectra concerning the missing species were added in a supplementary database. This new library was able to accurately predict unidentified species at first instance by MALDI-TOF MS, proving it is a powerful tool for the identification of environmental yeasts.

Keywords Yeast identification · MALDI-TOF MS · In-house database · Molecular biology · ITS region · Genetic sequencing

B. C. Agustini · G. A. da Silva (✉)
Laboratório de Microbiologia Aplicada, EMBRAPA Uva e Vinho, Bento Gonçalves, RS, Brazil
e-mail: gildo.almeida@embrapa.br

B. C. Agustini · T. M. B. Bonfim
Laboratório de Enzimologia e Tecnologia de Fermentações, Departamento de Farmácia, Universidade Federal do Paraná, Curitiba, PR, Brazil

L. P. Silva · C. Bloch Jr.
Laboratório de Espectrometria de Massa—LEM, EMBRAPA Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil

Introduction

Yeast species identification represents a major challenge in microbiology. The traditional process employs phenotypic methods to assess morphological, physiological, and biochemical characteristics. The assignment of yeast species has been established on morphological features of vegetative cells and sexual states, and on physiological responses on fermentation and assimilation tests using the taxonomic keys of Kreger-van Rij (1984), Barnett et al. (2000), and Lodder (1990). Application of morphological characteristics, including micro- and macroscopic, requires experienced microbiologists to interpret the results once different culture conditions as media and incubation time are sufficient to implicate in different outcomes (Velázquez et al. 2001). The use of nucleic acids sequencing to yeast systematics typically has pointed out discrepancies between their phenotype and genotype (Kurtzman and Robnett 1994; Cai et al. 1996). Moreover, in some cases, conventional identification does not present

sufficient discriminatory power to classify recently described species (Cendejas-Bueno et al. 2010).

Based on the confidence of molecular biology approaches, the restriction fragment length polymorphism analysis of PCR-amplified fragments (PCR-RFLP) from ribosomal internal transcribed spacer (ITS) region has been extensively employed for yeast differentiation (Combina et al. 2005; Guillamón et al. 1998; Wang and Liu 2013). Nevertheless, these methods are expensive and also time-consuming. Therefore, new technologies for accurate and rapid identification of yeast strains are essential to the different microbiological fields such as microbial culture collection and agribusiness.

A rapid and high-throughput identification method based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has been introduced in bacterial taxonomy (Demirev et al. 1999; Claydon et al. 1996; Holland et al. 1996) and also has been successfully applied in yeast and mold identification (Pan et al. 2011; Hendrickx et al. 2011; Marklein et al. 2009). MALDI-TOF MS has the ability to measure complex mixtures of proteins disclosing a unique fingerprint for each species, making it an important method for microorganism identification. Since the proteins detected are basically ribosomal ones, which are constitutively expressed at very high levels, this phenotypic technique is less influenced by expression variability (Wieser et al. 2012).

Besides MALDI-TOF MS has also been used successfully by several groups to differentiate yeast species isolated from clinical samples (van Veen et al. 2010; Marklein et al. 2009; Goyer et al. 2012; Qian et al. 2008; Stevenson et al. 2010; Dhiman et al. 2011), the data available for the differentiation of yeast isolates for industrial applications and environmental use are limited (Sherburn and Jenkins 2003; Moothoo-Padayachie et al. 2013; Usbeck et al. 2013; Blattel et al. 2013). It is worth noting that species identification relies on the comparison of the obtained spectrum from an unknown isolate with a previously available database. The outcome is generated based on similarities of average peak mass-to-charge (*m*/*z*) ratio, peak intensities, and peak frequencies (Bel et al. 2011). Once commercial databases comprise mostly clinically relevant microorganisms, studies focusing on environmental isolates are required to extend such databases. Hence, the proprietary MALDI Biotyper mass spectral database (Bruker Daltonics, Bremen, Germany) was selected for this study that aims to evaluate its applicability for taxonomic identifications of environmental isolates of a yeast collection. In addition, this study aims to develop a supplementary database in order to assess the performance of MALDI-TOF MS for the differentiation of other environmental yeast species improving the identification scope to biotechnological applications.

Material and methods

Yeast strains and culture conditions

The microorganisms were obtained from the Yeast Culture Collection (WDCM 1056) of the Centro Nacional de Pesquisa de Uva e Vinho (CNPUV, Bento Gonçalves, RS, Brazil) belonging to Brazilian Agricultural Research Corporation (EMBRAPA). All 845 strains were isolated from *Vitis vinifera* L. grapes from vineyards localized in Rio Grande do Sul and Pernambuco States.

The strains have been stored in cryogenic vials at −80 °C. For microorganisms reactivation, the cryogenic vials were defrosted at room temperature, 3 μL was transferred to Petri dishes containing must agar (da Silva 1996) and were cultured at 25 °C for 24–48 h. All species names were abbreviated according to Kreger-van Rij (1984).

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry-based identification

Sample preparation

For mass spectrometric analyses, the microorganisms were evaluated by cell extraction method. A small amount of biological material was taken from freshly grown colonies and transferred with a plastic loop into a polypropylene microtube containing 300 μL of ultra-pure water. Absolute ethanol (0.9 mL) was added, the contents of the microtube were mixed, and the microtubes were then centrifuged at 12,000× *g* for 2 min; the supernatant was discarded and the pellet was air-dried. Formic acid (70 %) and acetonitrile (ACN) were added in a 1:1 (*v*/*v*) ratio to the yeast pellet. The mixture was vortexed for 30 s and centrifuged for 2 min. An aliquot of 1 μL of the supernatant was transferred to a spot onto a 96-well stainless steel MALDI target plate. Applied cell supernatants were air-dried for 10 min following deposition of 1 μL of the α-cyano-4-hydroxycinnamic (CHCA, Sigma-Aldrich, São Paulo, Brazil) matrix prepared in an organic solvent mixture to a final concentration of 10 mg mL^{-1} in a 50:40:10 acetonitrile/water/3 % trifluoroacetic acid (TFA) solution that was overlaid and allowed to dry. Each sample was spotted in triplicate.

Mass spectra acquisition and data analysis

MALDI-TOF MS analysis of all strains was performed on a MicroFlex mass spectrometer (Bruker Daltonics, Bremen, Germany). The spectra were recorded in the linear positive mode at a laser frequency of 60 Hz within a mass range from *m*/*z* 2,000 to 20,000. For each spectrum, 240 laser shots in 40-shot steps from different positions of the target spot were collected and analyzed. The spectra were externally calibrated

by using *Escherichia coli* ribosomal proteins (Bruker Daltonics, Bremen, Germany).

To identify an unknown microorganism, the spectrum acquired was loaded with the MALDI Biotyper Software (Bruker Daltonics, Bremen, Germany) and analyzed by use of the standard pattern-matching algorithm, which compared the spectrum acquired with those present in the library. The fingerprints of unknown samples were compared to the fingerprints for all entries in the database, and the results were listed in a ranking table. The results of the pattern-matching process were expressed as log score values, which ranged from 0 to 3. Score values of >1.7 indicated identification beyond to the genus level, and score values of >2.0 indicated identification to the species level. Scores of <1.7 was interpreted as no identification.

FlexAnalysis Software (Bruker Daltonics, Bremen, Germany) was used for visual inspection and mass spectra processing such as smoothing, normalization, baseline subtraction, and peak picking.

Supplementary database/main spectra projection

Main spectra projections (MSPs) were created using four replicates of six separated colonies from the species to be included in the user-generated library. The main spectra were generated considering the 24 spectra obtained and are saved apart from the Biotyper database (Bruker Daltonics, Bremen, Germany). The addition of a given reference strain was made using the "MSP creation" function of the MALDI Biotyper software (version 3.0) comprising Bruker's default parameters (max. mass error of each single spectrum, 2,000; desired mass error for the MSP, 200; desired peak frequency minimum, 25 %; max. desired peak number for the MSP, 70). This supplementary database can be found on Mass Spectrometry Laboratory of Genetic Resources and Biotechnology Unit of EMBRAPA (Brasília, DF, Brazil). The MALDI-TOF mass spectra-based dendrogram was generated using the specific function present on Biotyper software.

The supplementary database was subsequently challenged against the spectra initially collected and not identified by Biotyper library.

Molecular identification

Identification using PCR-RFLP technique

Strains unsuccessfully identified by MALDI-TOF MS analysis were investigated through PCR-RFLP of ribosomal region spanning the ITS1, the 5.8S rRNA gene, and the ITS2. DNA extractions were carried out using the freeze–thawing process described by Silva et al. (2012). The primer pairs used were ITS1 and ITS4 described elsewhere (White et al. 1990). PCR was performed in 25 μL reaction volume containing 100 μM of each dNTP, 1× PCR buffer, 2.5 mM $MgCl_2$, 0.8 μM of each primer, 1.5 U Taq polymerase, and 1 μL DNA template. Amplifications were carried out in a PTC-100 thermocycler (M. J. Research, California, USA) using the following PCR program: 94 °C for 5 min followed by 25 cycles of 94 °C for 30 s, 60 °C for 45 s, 72 °C for 30 s, and a final step at 72 °C for 5 min. The endonucleases used were *Cfo*I, *Hae*III, and *Hin*fI. Digestion reactions contained 1.0 μL 10× digestion buffer (specific for each enzyme), 3 μL ultra-pure water, 1 μL restriction enzyme, and 5 μL PCR product. The temperature and incubation time followed the manufacturer recommendations for each enzyme. PCR products were resolved in 1 % agarose gel electrophoresis while restriction fragments were resolved in 3 % agarose gel electrophoresis. The gels were stained with ethidium bromide and the stained DNA was visualized under UV light on the Eagle Eye Image II (Stratagene, La Jolla, CA, USA). The fragments size were estimated by comparisons with a 100-bp DNA ladder.

Sequencing method

To assess the taxonomic identity of the resulting groups of ITS-RFLP used to create the Supplementary Database, the PCR product of the D1/D2 region of the large subunit of the 28S ribosomal RNA gene of at least one isolate of each group was sequenced. The sequencing primers were NL-1 and NL-4 (Kurtzman and Robnett 2003). Sequences were analyzed using Blast search at NCBI (http://www.ncbi.nlm.nih.gov/blast). These sequences were deposited on GenBank and received the following accession numbers KJ173770-KJ173777.

Results

Identifications by Biotyper database

We challenge the MALDI-TOF MS and MALDI Biotyper database (Bruker Daltonics, Bremen, Germany) to identify 845 yeast strains isolated from grape must. At first instance, only the database of the manufacturer was applied reaching 67.7 % of strains successfully identified when considering species and genera level (Table 1). The identified species comprised *Saccharomyces cerevisiae*, *Metschinikowia pulcherrima*, *Pichia guilliermondii*, *Hanseniaspora uvarum*, *Pichia galeiformis*, *Candida sorbosa*, and *Pichia kudriavzevii*.

All the species pointed out by MS approach were confirmed using molecular biology by submitting some strains of each species to ITS-RFLP methodology. This approach has recognized three strains of *Issatchenkia terricola* that were misidentified as *Pichia membranifaciens* by MS. Apart from this possibly erroneous identification, the other designations were accurate at species-level besides the occasionally low log scores obtained (<2.000). All the aforementioned species but

Table 1 Results of identification by MALDI-TOF MS using at first the database included in the software Biotyper (version 3.0.0.6) followed by the supplementary database

Species identified by:	Number of strains with results of log scores:			
	2.300–3.000	2.000–2.299	1.700–1.999	Total
Biotyper database				
S. cerevisiae	0	238	222	460
M. pulcherrima	0	0	5	5
P. guilliermondii	15	46	13	74
H. uvarum	4	8	0	12
P. galeiformis	0	1	3	4
C. sorbosa	0	0	5	5
C. krusei	1	5	3	9
Misidentification	0	0	3	3
Total	20	298	254	572
Supplementary database				
H. opuntiae	19	31	12	62
I. terricola	1	21	4	26
P. sporocuriosa	2	1	0	3
C. diversa	26	5	0	31
C. zemplinina	3	1	0	4
C. apicola	11	0	0	11
C. heveanensis	6	8	0	14
P. myanmarensis	102	17	3	122
Misidentifications	0	0	0	0
Total	170	84	19	273

The meanings of log score values were as follows: 2.300–3.000, highly probable species identification; 2.000–2.299, secure genus identification, probable species identification; 1.700–1.999, probable genus identification; and 0.000–1.699, no reliable identification

H. uvarum had some strains presenting results of log scores below 2.000, which is the manufacturer-recommended cutoff for species-level identification. The most representative case occurred with *S. cerevisiae* that presented almost 50 % of the analyzed strains with log score values confirming only the genus of the species (Table 1).

Table 2 shows the restriction patterns encountered for the species identified along the yeast collection studied. An intraspecific variability was observed to *P. galeiformis* and *Candida sorbosa* when comparing the fragments size with other studies (Guillamón et al. 1998; Bautista-Gallego et al. 2011).

Identifications by supplementary database

The strains unidentified by mass spectrometry database inspection in the first attempt were gathered according to the ITS-RFLP patterns, resulting in eight groups. The identity of the species was determined based on established literature (Esteve-Zarzoso et al. 1999; Guillamón et al. 1998; Pham et al. 2011), except for one pattern which had one representative strain directly submitted to the sequencing of the domains 1 and 2 of the 26S nuclear ribosomal gene due to the absent of a compatible RFLP pattern already showed in the literature. The BLAST search for this strain pointed to *Pichia myanmarensis* (KJ173777).

To assess the confidence among, and identity of the remaining seven groups, the sequence of the D1/D2 26S rDNA were determined for at least one example of each cohort. The D1/D2 26S rDNA is frequently used to discriminate among species, and empirical analyses suggests that isolates with <98 % homology in this region are likely different species (Kurtzman and Robnett 2003; Gayevskiy and Goddard 2012).

The groups formed were identified as *Hanseniaspora opuntiae* (KJ173776), *I. terricola* (KJ173770), *Candida diversa* (KJ173775), *Candida zemplinina* (KJ173771), *Cryptococcus heveanensis* (KJ173772), *Candida apicola* (KJ173773) and, the last pattern was identified as *Pichia sporocuriosa* or *Issatchenkia hanoiensis* (KJ173774).

One strain of each species absent in the manufacturer library was included in MALDI-TOF MS database as in-house library. To evaluate the reliability of the spectra generated, all the spectra collected in the first attempt of identification were submitted for a second round of Biotyper approach, now setting both manufacturer library and in-house library as the databases for microorganism search. All the strains not identified at first, after the inclusion of new species, had their best taxon matches in agreement with presumed species identification according to ITS-RFLP indications.

A dendrogram for a visual demonstration of the MALDI-TOF MS ability to differentiate among the included species was created using the Biotyper MSP dendrogram creation standard method (Fig. 1).

Discussion

Identifications by Biotyper database

MALDI-TOF MS has provided to be an accurate and rapid alternative method to traditional laboratory protocols aiming microorganisms' identification. Environmental yeast species and their identification may set a challenge when the type strains are not included in the MS database, leading to their misidentification as related species.

From the 845 yeast strains isolated from grape must, only 67.7 % of the strains were successfully identified at genera and species level using only the database provided by the manufacturer. Researches that analyzed clinical isolates of yeasts had reached better percentages of identification using only the Biotyper database as Marklein et al. (2009) obtained matches of 92.5 % and van Veen et al. (2010) reached 85.2 % for the clinical yeast species. The remaining 32.3 % of the strains tested did not result in any reliable matching (low log

Table 2 Size in base pair of the ITS-PCR products and the restriction fragments of the species present in the yeast collection analyzed

Species	ITS-PCR sizes (bp)	Size of restriction fragments (bp)			Reference for identification
		*Cfo*I	*Hae*III	*Hin*fI	
S. cerevisiae	880	370, 325, 140	310, 220, 170, 125	370, 360, 120	This study
	880	385, 365	320, 220, 180, 145	365, 155	Guillamón et al. (1998)
	850	380, 340	320, 225, 180, 145	360, 350, 120	Granchi et al. (1999)
	850	375, 325, 150	325, 230, 170, 125	375, 365, 110	Fernández-Espinar et al. (2000)
	880	380, 360, 140	340, 255, 175, 140	375, 130	Settanni et al. (2012)
	880	340, 320, 120	320, 240, 180, 140	390, 390, 100	Pham et al. (2011)
H. opuntiae	770	320, 315, 100	770	340, 180, 160	This study
	770	Not tested	690	340, 200, 170	Wang and Liu (2013)
	750	335, 115	750	370, 205, 175, 75	Settanni et al. (2012)
H. uvarum	770	320, 315, 100	770	340, 190, 160	This study
	760	320, 315, 105	760	360, 200, 180	Guillamón et al. (1998)
	770	Not tested	690	340, 200, 170	Wang and Liu (2013)
	750	335, 115	750	370, 205, 175, 75	Settanni et al. (2012)
	P. guilliermondii This study	625	300, 270, 40	400, 120, 80	310, 290
	625	300, 265, 60	400, 115, 90	320, 300	Dias et al. (2003)
625	300, 265, 60	400, 115, 90	320, 300	Esteve-Zarzoso et al. (1999)	
	P. myanmarensis	620	580, 40	540, 80	320, 180, 120
This study					
C. heveanensis	595	290, 285	370, 100, 90	250	This study
	595	278	358, 97	268	Lv et al. (2013)
P.kudriavzevii	500	200, 180, 75, 55	370, 90, 50	230, 150, 130	This study
	500	215, 190	400	230, 160	Settanni et al. (2012)
	500	200, 180, 70, 50	360, 90, 50	220, 150, 130	Pham et al. (2011)
P. galeiformis	490	230, 100, 70, 60	320, 140	200, 160, 100	This study
	450	250, 120, 80	325, 90	260, 220	Bautista-Gallego et al. (2011)
	470	230, 100	320, 80	260, 210	Villa-Carvajal et al. (2006)
	470	230, 100, 60	310, 90	230, 200	Rodríguez-Gómez et al. (2010)
C. apicola	490	200, 175	430	230, 120	This study
	490	220, 190, 100	400, 90	230, 130, 130	Frutos et al. (2004)
	500	205, 175	450	240, 125	Settanni et al. (2012)
	750	340, 310, 100	730	390, 195, 160	Esteve-Zarzoso et al. (1999)
C. sorbosa	480	215, 100, 80, 80	300, 90, 90	240, 110, 100	This study
	500	215, 100, 80	300, 90, 90	270, 120, 110	Sabate et al. (2002)
	610	575	610	325	Guillamón et al. (1998)
C. zemplinina	475	200, 100	475	240, 230	This study
	475	210, 110	475	235, 235	Settanni et al. (2012)
P. sporocuriosa	450	130, 95, 80, 70, 60	300, 140	240, 190	This study
	450	150, 120, 90	300, 150	250, 210	Hierro et al. (2006)
I. terricola	420	120, 90, 80, 75, 55	290, 120	220, 100, 100	This study
	416	120, 95, 78, 71, 58	290, 120	225, 100, 84	Granchi et al. (1999)
	420	125, 100, 90, 70	310, 110	225	Settanni et al. (2012)
	430	Not tested	290, 130	225, 105, 105	Wang and Liu (2013)
M. pulcherrima	390	210, 100, 95	285, 100	190, 180	This study
	400	200, 90	300, 100	200,180	Settanni et al. (2012)
	400	205, 100, 95	280, 100	200, 190	Sabate et al. (2002)
	390	210, 100	285, 100	200, 190	Guillamón et al. (1998)

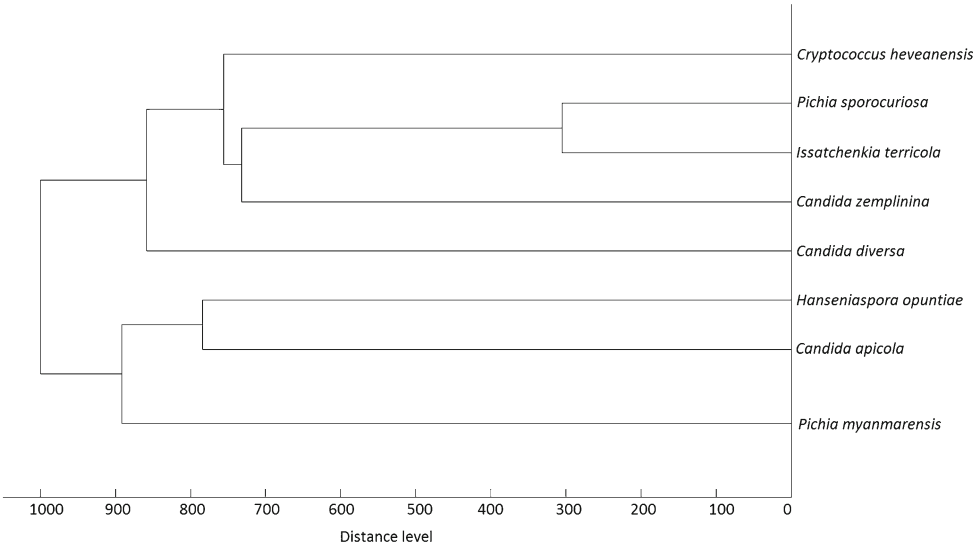

Fig. 1 MSP dendrogram for the eight yeast species included in the supplementary database

scores), suggesting the absence of appropriate reference spectra in the manufacturer database. Therefore, the lower percentage of identification in the present study highlights the preponderance presence of reference spectra towards species of clinical importance over environmental ones in the manufacturer database.

When confirming the species pointed out by Biotyper database, three strains were allocated as misidentifications. ITS-PCR product and its restriction fragments were compatible with *I. terricola*, but the absence of a reference spectrum for this microorganism lead to the designation as *P. membranifaciens*, which was considered by MALDI-TOF MS as the most similar match. This divergence was resolved by sequencing of D1/D2 26S region in favor of ITS-RFLP result. Thus, the ability of MALDI-TOF MS to differentiate both species after the addition of the missing spectra will be discussed below.

Bruker Daltonics company postulate log scores ≥2.000 to guarantee species-level identification using the MALDI Biotyper method. However, exception made for *H. uvarum*, all the designations made by Biotyper software were accurate besides the low log scores obtained (<2.000). According to Stevenson et al. (2010), all organisms identified at species level in their study had a spectral score of 1.8 or greater. Likewise, Pinto et al. (2011) pointed out that all the genus identification in their study also resolved isolates to the correct species designations.

Considering that supplementing the manufacturer database with in-house spectra increases the species-level identification (Marklein et al. 2009) combined with the evidenced ability of MS in discriminating strains (Moothoo-Padayachie et al. 2013), it is reasonable to infer that a log score below 2.000 is not sufficient to unequivocally affirm that the identification at species level is unreliable. It just reinforces that the ability to provide identifications depends on the number of entries per species in the manufacturer database. As the number of entries for the same species increases, better will be the representativeness of the species diversity due to possible variation in protein expression among strains (Pinto et al. 2011). The studies of Stevenson et al. (2010) and Dhiman et al. (2011) also indicated similar difficulties in obtaining log scores values ≥2.000 for yeast species of the genera *Aureobasidium*, *Candida*, and *Cryptococcus*, besides the existence of a reference spectra in the database selected. For these authors, such cases occurred due to inherent properties of the isolates or, as previously mentioned, to insufficient database entries to allow robust spectral matches.

This intraspecific variability is also observed in RFLP patterns during molecular biology identifications. When assuring the species identification pointed out by MS, some of them presented a distinct restriction profile from others reports in literature. For these cases, genetic sequencing was employed to confirm the taxonomy. The first case occurred with four isolates recognized like *Pichia manshurica* by Biotyper software when comparing Biotyper database after MALDI-TOF MS analyses. One of them was submitted to D1/D2 sequencing and BLAST search identified this strain as *P. manshurica* or *P. galeiformis*. Investigating both species, Ueda-Nishimura and Mikata (2001) have demonstrated that *P. galeiformis* is synonym of *P. manshurica* based on nuclear DNA composition, DNA/DNA hybridization, and comparison of 18S rRNA gene sequences. The four isolates of *P. galeiformis* identified showed distinct restriction pattern for *Hae*III and *Hin*fI compared with other studies in the literature (Rodríguez-Gómez et al. 2010; Villa-Carvajal et al. 2006; Muccilli et al. 2011). Discrepancies in restriction profiles among strains of a given species are not uncommon and had already been showed for several yeast species (Pham et al. 2011; Esteve-Zarzoso et al. 1999).

Strains of *Candida sorbosa* are encountered in the literature with amplicon sizes and RFLP pattern completely different among some studies. The isolates belonging to the present study are in accordance with previous report of Sabate et al. (2002) showing estimated amplicon sizes between 480 and 500 bp. However, as indicated in Table 2, the study of Guillamón et al. (1998) reported distinct ITS-RFLP profiles with ITS-PCR product size in 610 bp.

The species *M. pulcherrima*, *P. guilliermondii*, *H. uvarum*, and *P. kudriavzevii* demonstrated similar restriction profile as described by other authors (Guillamón et al. 1998; Granchi et al. 1999; Dias et al. 2003; Pham et al. 2011).

The last species identified by MALDI-TOF MS using only the manufacturer database was *S. cerevisiae.* The restriction profile of the tested strains of *S. cerevisiae* was also similar with previous literature reports (Fernández-Espinar et al. 2000; Settanni et al. 2012). Almost 50 % of the analyzed strains of *S. cerevisiae* presented log score values confirming only the genus of the species (Table 1). Possible explanations for the lower scores include an inefficient protein extraction for those samples or a large variability in protein expression within this species. Moreover, it has been noted that the spectra of this species suffer a great influence due to the colony age. Most studies reporting yeast identification by protein profiling using MALDI-TOF MS defines 48 h of culture growth before protein extraction (Marklein et al. 2009; Moothoo-Padayachie et al. 2013; Pinto et al. 2011). However, Goyer et al. (2012) have showed that either 48 or 72 h are suitable for the correct identification of yeast species. In the present study, we verified that the optimal growth period prior to protein extraction is strain dependent since for some of them higher scores (>2.0) were obtained with 72 h of growth while for others with just 48 h (data not shown). Our results are in agreement with Qian et al. (2008) that have also verified the strain-dependence for incubation condition. In general, the protocol established in this study considered colonies as young as 2 days old to initiate the extraction method, once an optimization for each strain would be impracticable due to the high number of samples included. Moreover, according to Wieser et al. (2012), colonies with more than 48 h of growth could produce weaker and less distinguishable peaks in the spectra due to probable ribosomal proteins hydrolysis.

Identifications by supplementary database

Based on the established literature, identifications by ITS-RFLP patterns were done for samples not identified by Biotyper database and, in sequence, the patterns identities were confirmed by D1/D2 sequencing.

As aforementioned, the ITS-RFLP profile for *P. myanmarensis* has not already been demonstrated; thus, the present study is the first report to set the resulting fragment sizes using endonuclease *Cfo*I, *Hae*III, and *Hin*fI (Table 2). This species is a novel taxon described by Nagatsuka et al. (2005) and shows high similarity with *Pichia anomala* concerning to 26S rDNA D1/D2 sequences. Likewise, the present study found *P. anomala* as the second best match (sharing 99 % nucleic acid sequence identity) for our strain.

Another cohort formed was recognized as *P. sporocuriosa* or *I. hanoiensis. P. sporocuriosa* and *I. hanoiensis* were described by Péter et al. (2000) and by Thanh et al. (2003), respectively. *Issatchenkia* and *Pichia* are recognized as closely related genera and a deep study towards phylogenetic relationships determined from multigene sequence analysis has showed that *P. sporocuriosa* and *I. hanoiensis* are identical species (Kurtzman et al. 2008). The authors explain that when *I. hanoiensis* were described, it was not recognized as conspecific with *P. sporocuriosa* because D1/D2 diagnostic sequences were not then available for the latter (Kurtzman et al. 2008). *P. sporocuriosa*/*I. hanoiensis* has been isolated from vines and wines in New Zealand (Gayevskiy and Goddard 2012), from the early stages of vinification with Tempranillo grapes in Spain (Hierro et al. 2006), from spontaneous wine fermentation of Catalanesca grapes in Italy (Di Maro et al. 2007), and of Castelao variety in Portugal (Baleiras Couto et al. 2005). The ITS-RFLP pattern for *P. sporocuriosa* has been demonstrated by Hierro et al. (2006) nonetheless the strains analyzed in the present study showed a different profile of cleavage with *Cfo*I and *Hin*fI (Table 2); hence, the genetic sequencing was also fundamental for this identification.

The intraspecific variability reflecting in distinct restriction profile was also observed for the strains of *Candida apicola.* The ones analyzed in this study were in accordance with previous report of Settanni et al. (2012) showing amplicon sizes between 490 and 500 bp. On the other hand, they are in disagreement with the study of Esteve-Zarzoso et al. (1999) that has described an ITS-PCR product size of 750 bp (Table 2).

Concerning the identified species *Cryptococcus heveanensis*, it has already been isolated from fermentation starters for glutinous rice wine (Lv et al. 2013), from leafs of strawberry (Debode et al. 2013), and from grape marcs (Bovo et al. 2009). And now, the present study reports this microorganism as an inhabitant of grape skin. Its restriction profile for *Cfo*I, *Hae*III, and *Hin*fI endonucleases has been showed by Lv et al. (2013) but the strain isolated from grape skin presented slight differences in the resulting restriction fragments (Table 2).

Tested strains of *Candida zemplinina* demonstrated similar restriction profile as showed by other authors (Settanni et al. 2012; Granchi et al. 1999; Wang and Liu 2013). Regarding the species *H. opuntiae*, its RFLP pattern was in agreement with previous reports (Wang and Liu 2013; Settanni et al. 2012); however, the two species of *Hanseniaspora* observed in the present study (*H. opuntiae* and *H. uvarum*) could not be

distinguished one from each other by ITS-RFLP using the three endonucleases chosen (Table 2). Wang and Liu (2013) solved this problem by employing endonucleases *Mbo*II and *Dde*I. In the present study, MS was sufficient to designate the correct species by protein profiling. The spectra of both species are visually unlike (Fig. 2). Exception made for the species of *Hanseniaspora*, all the remaining species obtained a pattern representative for its taxon employing endonucleases *Cfo*I, *Hae*III, and *Hin*fI.

A MSP consisting of an average of 24 spectra originated from 24 individual procedures of protein extraction were created for all eight species pointed out by genetic sequencing. The MSP represents the individual peak pattern including peak intensity distribution and peak frequency of the respective microorganism by extracting the typical peak information (Moothoo-Padayachie et al. 2013). After the construction of a supplementary database, all the strains were correctly identified proving the reliability of the new entries. The three strains primarily misidentified as *P. membranifaciens* by MALDI-TOF MS were included in this successful identification, and after the improvement of the database entries, they were corrected and identified as *I. terricola*. Misidentifications associated with the use of an incomplete database for yeast identification have been demonstrated by other studies (van Veen et al. 2010; Marklein et al. 2009).

Regarding the identification of *Candida diversa* by molecular biology, the strains isolated from grape berries surface failed in the amplification using universal primers ITS1 and ITS4 probably due to polymorphisms affecting the primers annealing. This unfeasibility to amplify the whole ITS1-5.8S-ITS2 region was described by Desnos-Ollivier et al. (2006) for four strains of *Madurella mycetomatis* suggesting the presence of mismatches in the target sequences. This fact reinforces the advantages of using MALDI-TOF MS for taxonomic identification since, after the addition of a reference strain in the database, this technique identified correctly all the strains that failed in the amplification of ITS region, apart from presenting a low consumable costs and minimum preparation time.

From the MSP dendrogram (Fig. 1), it was shown that the included strains could be clustered into two main clades. One of them contains the species *P. myanmarensis*, *H. opuntiae*, and *Candida apicola*. As previously mentioned, the genera *Pichia* and *Issatchenkia* are very similar; however, *P. myanmarensis* was positioned far from *P. sporocuriosa* and *I. terricola*. It is interesting to remind that the dendrogram generated from mass spectra does not represent phylogenetic relations. In this study, it can be assumed that species differentiation was possible up to a distance of 300 as can be seen for *P. sporocuriosa* and *I. terricola*. Marklein et al. (2009) have demonstrated the distinction of *Candida tropicalis* and *Candida albicans* up to a distance of 500. However, the potential of MALDI-TOF MS to identify microorganisms has been proved to reach distance level below 100 when concerning to strains differentiation (Christensen et al. 2012; Šedo et al. 2011; Moothoo-Padayachie et al. 2013).

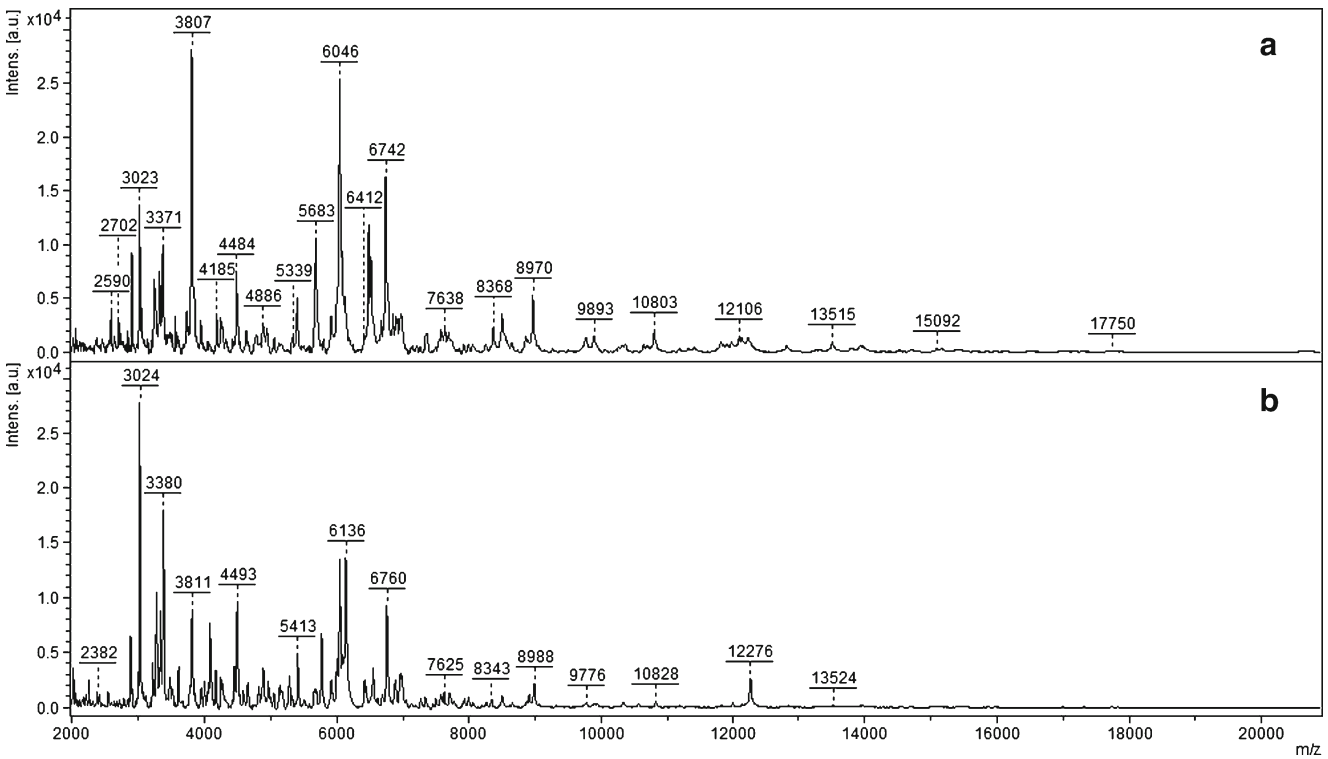

Fig. 2 Representative mass spectra of the comparison between *Hanseniaspora opuntiae* (**a**) and *Hanseniaspora uvarum* (**b**) yeast species

Comparing the results presented in Table 1, it is clear that in-house library favors the achievement of log scores above 2.300 (meaning highly probable species identification), besides the addition of the type-strains were set using only one strain of each species. This consideration shows the relatedness between the strains of the same species isolated from grape berries surface. Christensen et al. (2012) demonstrated the increase in log scores values after database extension, where the major part of the strains tested had achieved values between 2.300 and 3.000. Lau et al. (2013) developed a comprehensive database for identification of clinical molds and noted that the cutoff scores established by the manufacturer could be maintained without compromising sensitivity. The authors agree with this statement since 93.0 % of the identifications using in-house libraries achieved log score ≥2.0, against only 55.5 % using the manufacturer database (Table 1).

ITS-RFLP analyses, besides being a widespread taxonomic technique, can fail in the amplification of some strains or can be used only as an indicative of species identification due to similar profiles. On the other hand, the yeast database constructed to supplement Bruker MALDI Biotyper library have accurately predicted the species identity of all the 32.3 % of the strains not initially designated by the manufacturer database alone. Hence, it should be stressed that MALDI-TOF MS is a powerful tool for the identification of environmental yeast species, and that the continual expansion of the database is desirable to improve the scope of yeast identifications for non-clinical applications.

Acknowledgments The authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support.

References

Baleiras Couto MM, Reizinho RG, Duarte FL (2005) Partial 26S rDNA restriction analysis as a tool to characterise non-*Saccharomyces* yeasts present during red wine fermentations. Int J Food Microbiol 102:49–56

Barnett JA, Payne RW, Yarrow D (2000) Yeasts: characteristics and identification, 3rd edn. Cambridge University Press, Cambridge

Bautista-Gallego J, Rodríguez-Gómez F, Barrio E, Querol A, Garrido-Fernández A, Arroyo-López FN (2011) Exploring the yeast biodiversity of green table olive industrial fermentations for technological applications. Int J Food Microbiol 147:89–96

Bel AD, Wybo I, Vandoorslaer K, Rosseel P, Lauwers S, Piérard D (2011) Acceptance criteria for identification results of gram-negative rods by mass spectrometry. J Med Microbiol 60:684–686

Blattel V, Petri A, Rabenstein A, Kuever J, König H (2013) Differentiation of species of the genus *Saccharomyces* using biomolecular fingerprinting methods. Appl Microbiol Biotechnol 97: 4597–4606

Bovo B, Andrighetto C, Carlot M, Corich V, Lombardi A, Giacomini A (2009) Yeast population dynamics during pilot-scale storage of grape marcs for the production of Grappa, a traditional Italian alcoholic beverage. Int J Food Microbiol 129:221–228

Cai JP, Roberts IN, Collins MD (1996) Phylogenetic relationships among members of the ascomycetous yeast genera *Brettanomyces*, *Debaryomyces*, *Dekkera*, and *Kluyveromyces* deduced by small-subunit rRNA gene sequences. Int J Syst Bacteriol 46:542–549

Cendejas-Bueno E, Gomez-Lopez A, Mellado E, Rodriguez-Tudela JL, Cuenca-Estrella M (2010) Identification of pathogenic rare yeast species in clinical samples: comparison between phenotypical and molecular methods. J Clin Microbiol 48:1895–1899

Christensen JJ, Dargis R, Hammer M, Justesen US, Nielsen XC, Kemp M (2012) Matrix-assisted laser desorption ionization–time of flight mass spectrometry analysis of Gram-positive, catalase-negative cocci not belonging to the *Streptococcus* or *Enterococcus* genus and benefits of database extension. J Clin Microbiol 50:1787–1791

Claydon MA, Davey SN, Edwards-Jones V, Gordon DB (1996) The rapid identification of intact microorganisms using mass spectrometry. Nat Biotechnol 14:1584–1586

Combina M, Elía A, Mercado L, Catania C, Ganga A, Martinez C (2005) Dynamics of indigenous yeast populations during spontaneous fermentation of wines from Mendoza, Argentina. Int J Food Microbiol 99:237–243

da Silva GA (1996) The occurrence of killer, sensitive, and neutral yeasts in Brazilian Riesling Italico grape must and the effect of neutral strains on killing behaviour. Appl Microbiol Biotechnol 46:112–121

Debode J, van Hemelrijck W, Creemers P, Maes M (2013) Effect of fungicides on epiphytic yeasts associated with strawberry. Microbiol Open 2:482–491

Demirev PA, Ho Y-P, Ryzhov V, Fenselau C (1999) Microorganism identification by mass spectrometry and protein database searches. Anal Chem 71:2732–2738

Desnos-Ollivier M, Bretagne S, Dromer F, Lortholary O, Dannaoui E (2006) Molecular identification of black-grain mycetoma agents. J Clin Microbiol 44:3517–3523

Dhiman N, Hall L, Wohlfiel SL, Buckwalter SP, Wengenack NL (2011) Performance and cost analysis of matrix-assisted laser desorption ionization–time of flight mass spectrometry for routine identification of yeast. J Clin Microbiol 49:1614–1616

Di Maro E, Ercolini D, Coppola S (2007) Yeast dynamics during spontaneous wine fermentation of the Catalanesca grape. Int J Food Microbiol 117:201–210

Dias L, Dias S, Sancho T, Stender H, Querol A, Malfeito-Ferreira M, Loureiro V (2003) Identification of yeasts isolated from wine-related environments and capable of producing 4-ethylphenol. Food Microbiol 20:567–574

Esteve-Zarzoso B, Belloch C, Uruburu F, Querol A (1999) Identification of yeasts by RFLP analysis of the 5.85 rRNA gene and the two ribosomal interna transcribed spacers. Int J Syst Bacteriol 49:329–337

Fernández-Espinar MT, Esteve-Zarzoso B, Querol A, Barrio E (2000) RFLP analysis of the ribosomal internal transcribed spacers and the 5.8S rRNA gene region of the genus *Saccharomyces*: a fast method for species identification and the differentiation of flor yeasts. Anton van Leeuw 78:87–97

Frutos RD, Fernández-Espinar MT, Querol A (2004) Identification of species of the genus *Candida* by analysis of the 5.8S rRNA gene and the two ribosomal internal transcribed spacers. Anton van Leeuw 85:175–185

Gayevskiy V, Goddard MR (2012) Geographic delineations of yeast communities and populations associated with vines and wines in New Zealand. ISME J 6:1281–1290

Goyer M, Lucchi G, Ducoroy P, Vagner O, Bonnin A, Dalle F (2012) Optimization of the preanalytical steps of matrix-assisted laser desorption ionization–time of flight mass spectrometry identification provides a flexible and efficient tool for identification of clinical yeast isolates in medical laboratories. J Clin Microbiol 50:3066–3068

Granchi L, Bosco M, Messini A, Vincenzini M (1999) Rapid detection and quantification of yeast species during spontaneous wine fermentation by PCR-RFLP analysis of the rDNA ITS region. J Appl Microbiol 87:949–956

Guillamón JM, Sabaté J, Barrio E, Cano J, Querol A (1998) Rapid identification of wine yeast species based on RFLP analysis of the ribosomal internal transcribed spacer (ITS) region. Arch Microbiol 169:387–392

Hendrickx M, Goffinet J-S, Swinne D, Detandt M (2011) Screening of strains of the *Candida parapsilosis* group of the BCCM/IHEM collection by MALDI-TOF MS. Diagn Microbiol Infect Dis 70:544–548

Hierro N, González A, Mas A, Guillamón JM (2006) Diversity and evolution of non-*Saccharomyces* yeast populations during wine fermentation: effect of grape ripeness and cold maceration. FEMS Yeast Res 6:102–111

Holland RD, Wilkes JG, Rafii F, Sutherland JB, Persons CC, Voorhees KJ, Lay JO (1996) 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 10:1227–1232

Kreger-van Rij NJW (1984) The yeasts: a taxonomic study, 3rd edn. Elsevier, Amsterdam

Kurtzman CP, Robnett CJ (1994) Synonymy of the yeast genera *Wingea* and *Debaryomyces*. Anton van Leeuw 66:337–342

Kurtzman CP, Robnett CJ (2003) Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast Res 3:417–432

Kurtzman CP, Robnett CJ, Basehoar-Powers E (2008) Phylogenetic relationships among species of *Pichia*, *Issatchenkia* and *Williopsis* determined from multigene sequence analysis, and the proposal of *Barnettozyma* gen.nov., *Lindnera* gen.nov. and *Wickerhamomyces* gen.nov. FEMS Yeast Res 8:939–954

Lau AF, Drake SK, Calhoun LB, Henderson CM, Zelazny AM (2013) Development of a clinically comprehensive database and a simple procedure for identification of molds from solid media by matrix-assisted laser desorption ionization–time of flight mass spectrometry. J Clin Microbiol 51:828–834

Lodder J (1990) Criteria and methods used in classification in the yeast: a taxonomic study. North-Holland, Amsterdam

Lv X-C, Huang X-L, Zhang W, Rao P-F, Ni L (2013) Yeast diversity of traditional alcohol fermentation starters for Hong Qu glutinous rice wine brewing, revealed by culture-dependent and culture-independent methods. Food Control 34:183–190

Marklein G, Josten M, Klanke U, Müller E, Horré R, Maier T, Wenzel T, Kostrzewa M, Bierbaum G, Hoerauf A, Sahl H-G (2009) Matrix-assisted laser desorption ionization–time of flight mass spectrometry for fast and reliable identification of clinical yeast isolates. J Clin Microbiol 47:2912–2917

Moothoo-Padayachie A, Kandappa HR, Krishna SBN, Maier T, Govender P (2013) Biotyping *Saccharomyces cerevisiae* strains using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Eur Food Res Technol 236:351–364

Muccilli S, Caggia C, Randazzo CL, Restuccia C (2011) Yeast dynamics during the fermentation of brined green olives treated in the field with kaolin and Bordeaux mixture to control the olive fruit fly. Int J Food Microbiol 148:15–22

Nagatsuka Y, Kawasaki H, Seki T (2005) *Pichia myanmarensis* sp. nov., a novel cation-tolerant yeast isolated from palm sugar in Myanmar. Int J Syst Evol Microbiol 55:1379–1382

Pan Y-L, Chow N-H, Chang TC, Chang H-C (2011) Identification of lethal *Aspergillus* at early growth stages based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Diagn Microbiol Infect Dis 70:344–354

Péter G, Tornai-Lehoczki J, Dlauchy D, Vitányi G (2000) *Pichia sporocuriosa* sp. nov., a new yeast isolated from rambutan. Anton van Leeuw 77:37–42

Pham T, Wimalasena T, Box WG, Koivuranta K, Storgårds E, Smart KA, Gibson BR (2011) Evaluation of ITS PCR and RFLP for differentiation and identification of brewing yeast and brewery 'wild' yeast contaminants. J I Brewing 117:556–568

Pinto A, Halliday C, Zahra M, van Hal S, Olma T, Maszewska K, Iredell JR, Meyer W, Chen SC (2011) Matrix-assisted laser desorption ionization–time of flight mass spectrometry identification of yeasts is contingent on robust reference spectra. PLoS ONE 6:e25712

Qian J, Cutler J, Cole R, Cai Y (2008) MALDI-TOF mass signatures for differentiation of yeast species, strain grouping and monitoring of morphogenesis markers. Anal Bioanal Chem 392:439–449

Rodríguez-Gómez F, Arroyo-López FN, López-López A, Bautista-Gallego J, Garrido-Fernández A (2010) Lipolytic activity of the yeast species associated with the fermentation/storage phase of ripe olive processing. Food Microbiol 27:604–612

Sabate J, Cano J, Esteve-Zarzoso B, Guillamón JM (2002) Isolation and identification of yeasts associated with vineyard and winery by RFLP analysis of ribosomal genes and mitochondrial DNA. Microbiol Res 157:267–274

Šedo O, Voráč A, Zdráhal Z (2011) Optimization of mass spectral features in MALDI-TOF MS profiling of *Acinetobacter* species. Syst Appl Microbiol 34:30–34

Settanni L, Sannino C, Francesca N, Guarcello R, Moschetti G (2012) Yeast ecology of vineyards within Marsala wine area (western Sicily) in two consecutive vintages and selection of autochthonous *Saccharomyces cerevisiae* strains. J Bios Bioeng 114:606–614

Sherburn RE, Jenkins RO (2003) A novel and rapid approach to yeast differentiation using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Spectroscopy 17:31–38

Silva GA, Bernardi TL, Schaker PDC, Menegotto M, Valente P (2012) Rapid yeast DNA extraction by boiling and freeze–thawing without using chemical reagents and DNA purification. Braz Arch Biol Techn 55:319–327

Stevenson LG, Drake SK, Shea YR, Zelazny AM, Murray PR (2010) Evaluation of matrix-assisted laser desorption ionization–time of flight mass spectrometry for identification of clinically important yeast species. J Clin Microbiol 48:3482–3486

Thanh VN, Hai DA, Lachance MA (2003) *Issatchenkia hanoiensis*, a new yeast species isolated from frass of the litchi fruit borer *Conopomorpha cramerella* Snellen. FEMS Yeast Res 4:113–117

Ueda-Nishimura K, Mikata K (2001) Reclassification of *Pichia scaptomyzae* and *Pichia galeiformis*. Anton van Leeuw 79:371–375

Usbeck JC, Kern CC, Vogel RF, Behr J (2013) Optimization of experimental and modelling parameters for the differentiation of beverage spoiling yeasts by matrix-assisted-laser-desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS) in response to varying growth conditions. Food Microbiol 36:379–387

van Veen SQ, Claas EC, Kuijper EJ (2010) High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization–time of flight mass spectrometry in conventional medical microbiology laboratories. J Clin Microbiol 48:900–907

Velázquez E, Cruz-Sánchez JM, Rivas-Palá T, Zurdo-Piñeiro JL, Mateos PF, Monte E, Martínez-Molina E, Chordi A (2001) YeastIdent-Food/ProleFood, a new system for the identification of food yeasts based on physiological and biochemical tests. Food Microbiol 18:637–646

Villa-Carvajal M, Querol A, Belloch C (2006) Identification of species in the genus *Pichia* by restriction of the internal transcribed spacers (ITS1 and ITS2) and the 5.8S ribosomal DNA gene. Anton van Leeuw 90:171–181

Wang C, Liu Y (2013) Dynamic study of yeast species and *Saccharomyces cerevisiae* strains during the spontaneous fermentations of Muscat blanc in Jingyang, China. Food Microbiol 33:172–177

White TJ, Bruns TD, Lee E, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp 315–322

Wieser A, Schneider L, Jung J, Schubert S (2012) MALDI-TOF MS in microbiological diagnostics—identification of microorganisms and beyond (mini review). Appl Microbiol Biotechnol 93:965–974