Yeast
Yeast 2016; **33**: 433–449.
Published online 13 May 2016 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/yea.3162

ISSY32 Special Issue

Proteomic analysis of *Rhodotorula mucilaginosa*: dealing with the issues of a non-conventional yeast

Maria Filippa Addis[1†], Alessandro Tanca[1†], Sara Landolfo[2], Marcello Abbondio[1], Raffaela Cutzu[2], Grazia Biosa[1], Daniela Pagnozzi[1], Sergio Uzzau[1,3] and Ilaria Mannazzu[2*]

[1]*Porto Conte Ricerche, Tramariglio Alghero, Italy*
[2]*Dipartimento di Agraria, Università di Sassari, Italy*
[3]*Dipartimento di Scienze Biomediche, Università di Sassari, Italy*

**Correspondence to:*
I. Mannazzu, Dipartimento di Agraria, Università di Sassari, Viale Italia 39, 07100 Sassari, Italy.
E-mail: imannazzu@uniss.it

†These authors contributed equally to this study.

Received: 30 December 2015
Accepted: 9 March 2016

Abstract

Red yeasts ascribed to the species *Rhodotorula mucilaginosa* are gaining increasing attention, due to their numerous biotechnological applications, spanning carotenoid production, liquid bioremediation, heavy metal biotransformation and antifungal and plant growth-promoting actions, but also for their role as opportunistic pathogens. Nevertheless, their characterization at the 'omic' level is still scarce. Here, we applied different proteomic workflows to *R. mucilaginosa* with the aim of assessing their potential in generating information on proteins and functions of biotechnological interest, with a particular focus on the carotenogenic pathway. After optimization of protein extraction, we tested several gel-based (including 2D-DIGE) and gel-free sample preparation techniques, followed by tandem mass spectrometry analysis. Contextually, we evaluated different bioinformatic strategies for protein identification and interpretation of the biological significance of the dataset. When 2D-DIGE analysis was applied, not all spots returned a unambiguous identification and no carotenogenic enzymes were identified, even upon the application of different database search strategies. Then, the application of shotgun proteomic workflows with varying levels of sensitivity provided a picture of the information depth that can be reached with different analytical resources, and resulted in a plethora of information on *R. mucilaginosa* metabolism. However, also in these cases no proteins related to the carotenogenic pathway were identified, thus indicating that further improvements in sequence databases and functional annotations are strictly needed for increasing the outcome of proteomic analysis of this and other non-conventional yeasts. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: **2D-DIGE; shotgun proteomics; protein extraction; red yeast; carotenoid**

Introduction

Yeasts ascribed to the genus *Rhodotorula* Harrison are basidiomycetes (subphylum Pucciniomycotina) occurring in terrestrial, aquatic and marine habitats. These yeasts are natural inhabitants of the phylloplane and of decaying plants, but can also be isolated from air, soil, food, stool and human skin (Fell and Stratzell-Tallman, 1998). Their strong oxidative metabolism enables the degradation of recalcitrant substrates, organochemicals and industrial wastes (Cheirsilp *et al*., 2011; Johnson, 2013; Taskin, 2013). They are also naturally capable of bioconverting a variety of by-products of the agrifood industry into added value primary and secondary metabolites and, based on that, have been proposed as a source of pigments and metabolites of interest in the food industry (Hernández-Almanza *et al*., 2014), as oil producers for biofuel application (Galafassi *et al*., 2012; Li *et al*., 2010; Tampitak

et al., 2015) and as enzyme producers (Canli *et al*., 2011; Taskin, 2013). Moreover, these yeasts show antimicrobial activity against pathogenic fungi involved in postharvest diseases of fruit and vegetables (Li *et al*., 2011; Zhang *et al*., 2008, 2014).

Among the different species ascribed to the genus, *Rhodotorula mucilaginosa* is highly heterogeneous at the genetic and phenotypic levels and presents a number of synonyms (Libkind *et al*., 2008). Yeasts ascribed to this species are the object of increasing interest, due to their biotechnological potential. Accordingly, a number of strains of *R. mucilaginosa* have been explored for the commercial production of carotenoids, mainly β-carotene, torulene and torularhodin (Aksu and Eren, 2005; Maldonade *et al*., 2012), for liquid bioremediation processes (Jarboui *et al*., 2012), for heavy metal biotransformation (Rajpert *et al*., 2013) and for their antifungal and plant growth-promoting actions (Ignatova *et al*., 2015). In addition, there is also interest concerning their role as emergent opportunistic pathogens in immunocompromised individuals (Wirth and Goldani, 2012).

Recently, the draft genome sequence of *R. mucilaginosa* has been released (Deligios *et al*., 2015). However, although our knowledge on red yeasts is expected to increase in the near future, due to the renewed interest on their biotechnological potential and the recent development of molecular and -omic tools (Mannazzu *et al*., 2015), this species is still poorly characterized at the genomic and proteomic level. Concerning proteome characterization, different approaches were applied to uncover the salt stress response (Lahav *et al*., 2004) and the mechanisms underlying copper resistance in *R. mucilaginosa* (Irazusta *et al*., 2012) in view of its possible uses in bioremediation. Moreover, a number of studies regarding the optimization of carotenoid production by *R. mucilaginosa* have been published, but further information on the location and the tight and complex regulation of the carotenogenic pathway is needed to convert these non-conventional yeasts in promising biocatalysts.

Here, with the aim of contributing to the development of molecular tools useful for the identification of proteins and functions of potential biotechnological interest in the non-conventional yeast *R. mucilaginosa*, we applied different proteomic approaches, from 2D-DIGE to an array of other gel-based and non-gel-based proteomic analysis workflows, and evaluated their output in a critical and comparative manner. Moreover, we evaluated the impact of using different sequence databases on improvements of the proteomic analysis outcome. Then, in order to assess their potential for the dissection of pathways of biotechnological interest, we investigated the level of proteomic information generated on the carotenogenic pathway with the different workflows.

Materials and methods

Strains, culture conditions and carotenoid analysis

The strains utilized were: C2.5t1, previously identified as *R. glutinis* and recently ascribed to the species *R. mucilaginosa* (Deligios *et al*., 2015), deposited at the Yeast Collection of Dipartimento di Scienze della Vita e dell'Ambiente (DiSVA), Universita` Politecnica delle Marche (Ancona, Italy); and 400A15 and 200A6, primary mutants of C2.5t1 deposited at the Culture Collection of the Dipartimento di Agraria, Università degli Studi di Sassari (Sassari, Italy). Mutants were obtained by UV mutagenesis, as described by Cutzu *et al*. (2013b). The yeasts were maintained on YEPD (glucose 2%, yeast extract 1%, bacto-peptone 2%, agar 2%) at 4 °C for short-term conservation, and on YEPD with added 20% glycerol at –80 °C for long-term storage. The yeasts were grown in 250 ml baffled flasks containing 50 ml YEPGLY (glycerol 2%, yeast extract 1%, bacto-peptone 2%) under shaking conditions (180 rpm) at 30 °C, and collected after 16, 40 and 72 h (Cutzu *et al*., 2013a, 2013b). Carotenoid extraction and quantification was carried out as described by Cutzu *et al*. (2013b). Unless otherwise stated, biological replicates were obtained from at least three independent cultures.

Protein extraction and quantitation

Protein extraction from *R. mucilaginosa* cells was carried out according to three different methods (thiourea–urea–CHAPS buffer with mechanical disruption, TUC–MD; thiourea–urea–CHAPS

DOI: 10.1002/yea

buffer with mechanical disruption plus sonication, TUC–MD + S; SDS-based buffer with thermal shock plus mechanical disruption, SDS–TS + MD), as detailed below. Initially, 1×10^9 cells were resuspended in two-dimensional (2D) Protein Extraction Buffer III (GE Healthcare, Little Chalfont, UK) (TUC–MD and TUC–MD + S) or in a preheated (95 °C) buffered solution containing 20 mM Tris–HCl, pH 8.8, 1% sodium dodecyl sulphate (SDS) (SDS–TS + MD). In all cases, protease inhibitors were added (SIGMAFAST™ Protease Inhibitor Tablets, Sigma-Aldrich, St. Louis, MO, USA), according to the manufacturer's instructions. Then the samples were directly subjected to mechanical homogenization (TUC–MD), sonicated in a Transsonic Digital ultrasonic bath (Elma Electronic, Wetzikon, Switzerland) and then subjected to mechanical homogenization (TUC–MD + S) or incubated at 95 °C for 20 min at 500 rpm in a Thermomixer Comfort (Eppendorf, Hamburg, Germany), followed by a combination of freeze–thawing cycles, as described previously (Tanca *et al*., 2014b) (SDS–TS + MD). For the mechanical homogenization procedure, a stainless steel bead (5 mm diameter; Qiagen, Hilden, Germany) was added to each sample, after which the samples were subjected to bead beating for 10 min at 30 cycles/s in a TissueLyser mechanical homogenizer (Qiagen). The samples were finally centrifuged at 14 000 × *g* for 10 min at 4 °C and the whole supernatants were collected. Standard colourimetric and fluorimetric protein quantification assays could not be applied, due to the strong interference caused by the yeast pigments. Protein concentration was therefore estimated by whole-lane densitometry, using QuantityOne software (Bio-Rad, Hercules, CA, USA) after electrophoretic separation through an Any kD Mini-PROTEAN TGX Gel (Bio-Rad) and gel staining with SimplyBlue SafeStain (Invitrogen, Carlsbad, CA, USA).

2D-PAGE

Proteins extracted from three independent 72 h cultures of *R. mucilaginosa* strain C2.5 t1 using the SDS–TS + MD method were analysed by 2D-PAGE, according to two different protocols.

In the first case, proteins were sequentially diluted 1:10 in 2D Protein Extraction Buffer III (GE Healthcare) and labelled with CyDye Fluor 3 (GE Healthcare), according to the minimal labelling protocol provided by the manufacturer. IPG buffer (pH 3–11 NL; GE Healthcare) was added to a final concentration of 1% and DeStreak Rehydration Solution (GE Healthcare) was added to a total volume of 200 µl. First-dimension isoelectro focusing (IEF) was carried out using 3–10 NL 11 cm IPG strips (BioRad, Hercules, CA, USA), which were passively rehydrated overnight at room temperature and focused on IPGphor equipped with the Ettan™IPGphor3™ loading manifold (GE Healthcare) at 20 °C. After focusing, the strips were equilibrated, reduced and alkylated by sequential incubation in 2% w/v dithiotreitol (DTT) and 2.5% w/v iodoacetamide (IAM), both in 50 mM Tris–HCl, pH 8.8, 6 M urea, 30% v/v glycerol and 2% w/v SDS for 15 min. Second-dimension SDS–polyacrylamide gel electrophoresis (SDS–PAGE) was conducted in a Criterion Dodeca Cell (Bio-Rad) with 4–20% and AnyKD Criterion gels (Bio-Rad), as described previously (Addis *et al*., 2012).

In the second case, proteins underwent precipitation using the 2D-Clean up Kit (GE Healthcare), according to manufacturer's instructions, and the obtained pellets were resuspended in 2D Protein Extraction Buffer V (GE Healthcare) plus 10 mM Tris–HCl, pH 8.8, to facilitate solubilization. IPG buffer (pH 3–11 NL; GE Healthcare) was added to a final concentration of 1%, and DeStreak Rehydration Solution (GE Healthcare) was added to a total volume of 450 µl. First-dimension IEF was carried out using 3–11 NL 24 cm IPG strips (GE Healthcare), which were passively rehydrated overnight at room temperature and focused on IPGphor equipped with the Ettan™IPGphor3™ loading manifold (GE Healthcare) at 20 °C. After focusing, the strips were equilibrated, reduced and alkylated as above. The second-dimension SDS–PAGE was performed as described previously (Tanca *et al*., 2013c), using 12.5% polyacrylamide gels in an Ettan DALT Twelve electrophoresis system (GE Healthcare). Gel images were acquired on a Typhoon Trio + image scanner (GE Healthcare) at a resolution of 100 µm.

2D-DIGE

2D-DIGE experiments were performed according to the second sample preparation protocol described in the previous section. Concerning

DOI: 10.1002/yea

protein-labelling schemes, in the first experiment (growth curve analysis) samples collected at three sequential points of the growth curve (16, 40 and 72 h) were labelled with CyDye 2, 3 and 5, respectively; in the second experiment (mutant strains comparison) proteins were labelled with CyDye DIGE Fluors 3–5 (GE Healthcare), according to the minimal labelling protocol provided by the manufacturer, and, in parallel, a mixture of all samples was labelled with CyDye DIGE Fluor 2 and employed as a pooled internal standard. Gel images were acquired on a Typhoon Trio + image scanner (GE Healthcare) at 100 µm resolution and exported to the Batch Processor and Differential In-gel Analysis (DIA) modules of Decyder 2D v. 7.0 software (GE Healthcare) for statistical analysis. The results were compared and statistically evaluated by one-way analysis of variance (ANOVA) using the biological variation analysis (BVA) module, applying the false-discovery rate (FDR) to minimize the number of false-positive results. Protein spots with statistically significant variation ($p < 0.05$) and average ratio > 1.5 or < -1.5 were selected as differentially expressed.

Separation, digestion, MS analysis and identification of differential protein spots

Preparative 2D gels were generated in order to cut and identify differential protein spots selected upon DIGE analysis. Specifically, 600 µg C2.5 t1 and 400A15 protein extract pools were loaded into 3–11 NL 24 cm strips (GE Healthcare) and 12.5% polyacrylamide gels. The same protocol used for 2D-DIGE was followed, except for CyDye labelling. The gels were subjected to Coomassie R-250 staining (Westermeier, 2006), digitalized by scanning with an ImageScanner III (GE Healthcare) and matched to the respective 2D-DIGE gel image, in order to track the spots to be excised for protein identification. Spots of interest were manually excised from the gels, destained and subjected to overnight tryptic digestion, as described previously (Tanca *et al*., 2011).

LC–MS/MS analysis of peptides was performed on a XCT Ultra 6340 ion trap equipped with a 1200 HPLC system and a chip cube (Agilent Technologies, Palo Alto, CA, USA), according to an established procedure (Ghisaura *et al*., 2014). Peak lists generated from MS/MS spectra were analysed using Proteome Discoverer v. 1.4.1.14 (Thermo Scientific), using Sequest-HT as search engine for peptide identification, with the following parameters: trypsin as enzyme; maximum of two missed cleavage sites; 250 ppm precursor mass tolerance; 0.5 Da fragment mass tolerance; cysteine carbamidomethylation as static modifications; methionine oxidation as dynamic modification. Peptide identifications were filtered according to a 1% FDR threshold, based on the Target-Decoy Peptide Validator tool provided by Proteome Discoverer. Peptide and protein groupings according to Proteome Discoverer's algorithms were allowed, applying a strict maximum parsimony principle.

Three alternative sequence databases were used for peptide identification from gel spots: S-DB, comprising all UniProtKB (v. 2015_02) sequences belonging to the order Sporidiobolales (13 005 sequences); C-DB, comprising all UniProtKB (v. 2015_02) sequences related to the carotenoid biosynthesis pathway (starting from acetyl-coenzyme A acetyltransferase to lycopene cyclase), irrespective of the organism of provenience, with clustering at 90% homology (243 146 sequences); and G-DB, comprising all sequences generated upon genome sequencing of the C2.5t1 strain, as described in Deligios *et al*. (2015) (6412 sequences, of which 4352 associated to a UniProt Accession No. upon blastp alignment).

Gel-based fractionation and ion trap LC–MS/MS analysis of protein extracts (EGIT)

Samples (30 µg) of each protein extract were separated in an AnyKD TGX gel (Bio-Rad) and stained with SimplyBlue SafeStain (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. Then each lane was fractionated into 26 gel slices, which were destained, reduced, carbamidomethylated and trypsin-digested, as described previously (Pisanu *et al*., 2013; Tanca *et al*., 2012). LC–MS/MS analysis of peptides was performed on a XCT Ultra 6340 ion trap equipped with a 1200 HPLC system and a chip cube (Agilent), as above. Peak lists generated from MS/MS spectra were analysed by Proteome Discoverer, using Sequest-HT as the search engine for peptide identification, with the same parameters described in the previous section. Peptide significance was validated at the peptide level, based on Percolator q values ($q < 0.01$). The 'merged' sequence database employed here (262 527

DOI: 10.1002/yea

sequences) was generated by merging the three sequence databases described above.

FASP and LTQ-Orbitrap Velos LC–MS/MS analysis of protein extracts (EFOV)

Samples (30 µg) of protein extract were diluted to 200 µL with 8 M urea, 100 mM Tris–HCl, pH 8.8, loaded into a Microcon Ultracel YM-30 filtration device (Merck Millipore, Billerica, MA, USA) and then processed according to the 'FASP II' protocol (Wiśniewski *et al*., 2009), with minor modifications (Tanca *et al*., 2013a). Briefly, samples were subjected to repetitive washings by filter centrifugations with buffers, DTT and IAM, followed by overnight on-filter digestion with trypsin, peptide elution using acetonitrile (ACN) and formic acid, an additional step with Ultrafree MC-GV centrifugal filters (Merck Millipore), drying, and final reconstitution of the peptide mixture in 0.2% formic acid. Peptide concentration was estimated by measuring absorbance at 280 nm with a NanoDrop 2000 spectrophotometer (Thermo Scientific, San Jose, CA, USA), using dilutions of the MassPREP *Escherichia coli* Digest Standard (Waters, Milford, MA, USA) to generate a calibration curve, as illustrated elsewhere (Tanca *et al*., 2014a). Peptide mixtures (4 µg) were analysed using an LTQ-Orbitrap Velos interfaced with an UltiMate 3000 RSLCnano LC system (both from Thermo Scientific), using a 485 min gradient for peptide separation, as described previously (Tanca *et al*., 2014b). Peak lists generated from MS/MS spectra were analysed by Proteome Discoverer, using Sequest-HT as the search engine for peptide identification with the same parameters illustrated above, except for precursor mass tolerance (10 ppm) and fragment mass tolerance (0.02 Da). Peptide significance was validated at the peptide level, based on Percolator q values ($q < 0.01$). A 'merged' sequence database was employed as described above.

Gel-based fractionation and Q-Exactive LC–MS/MS analysis of residual pellets (PGQE)

Samples (10 mg) of residual extraction pellets (see section 'Protein extraction and quantitation') were incubated with 50 µl Laemmli buffer (Laemmli, 1970) at 95 °C for 15 min and partially separated in an AnyKD TGX gel (Bio-Rad) for 5 min (Paulo *et al*., 2013); then, five slices/sample were cut and destained, reduced, carbamidomethylated and trypsin-digested, as above. Peptides were analysed using a Q-Exactive mass spectrometer interfaced with an UltiMate 3000 RSLCnano LC system (both from Thermo Scientific). Peptide LC separation was carried out using a 485 min gradient, as for the EFOV method and in a previous study (Tanca *et al*., 2014b). MS data were acquired using a data-dependent top 10 method, dynamically choosing the most abundant precursor ions from the survey scan, under direct control of Xcalibur software (v. 1.0.2.65 SP2), where a full-scan spectrum (300–1700 m/z) was followed by tandem mass spectrometry (MS/MS). The instrument was operated in positive mode, with a spray voltage of 1.8 kV and a capillary temperature of 275 °C. Survey and MS/MS scans were performed in the Orbitrap with a resolution of 70 000 and 17 500 at 200 m/z, respectively. The automatic gain control was set to 1 000 000 ions and the lock mass option was enabled on a protonated polydimethylcyclosiloxane background ion as internal recalibration for accurate mass measurements. The dynamic exclusion was set to 30 s. Higher-energy collisional dissociation (HCD), performed at the far side of the C-trap, was used as the fragmentation method by applying a 25 eV value for normalized collision energy and an isolation width of m/z 2.0. Nitrogen was used as the collision gas. Peptide identification tools, database and parameters, as well as the peptide validation method, were as described for the EFOV approach.

Shotgun proteomic data analysis

The normalized spectral abundance factor (NSAF) was calculated as described elsewhere (Tanca *et al*., 2014a; Zybailov *et al*., 2006) and used to estimate peptide abundance. The relative abundance of a feature (protein or functional category) was calculated by summing the NSAF values of all peptides matched to that given feature. The NSAF log ratio was calculated as previously described (Tanca *et al*., 2012), using 2 as the correction factor, and employed to estimate the extent of differential abundance. Gene Ontology categories were retrieved from the UniProt website (http://www.uniprot.org; 2015). KEGG orthology groups (KOGs) information was gathered using KAAS

DOI: 10.1002/yea

(http://www.genome.jp/tools/kaas) (Moriya *et al*., 2007). The number of transmembrane domains within protein sequences was predicted using the TMHMM Server (v. 2.0; http://www.cbs.dtu.dk/services/TMHMM) (Krogh *et al*., 2001). The interactive Pathways Explorer (iPath v.2, http://pathways.embl.de) was used to map proteins into metabolic pathways (Yamada *et al*., 2011). Data were parsed using in-house scripts, and graphs were generated using Microsoft Excel and Venn Diagram Plotter (http://omics.pnl.gov/software/venn-diagram-plotter).

Results

Optimization of sample preparation for protein electrophoresis

The first aim of our study was to devise an efficient protocol for extraction and solubilization of *R. mucilaginosa* proteins, able to combine high yields with compatibility with 2D-PAGE analysis. Therefore, a classical 2DE-compatible buffer (TUC) was initially used for solubilizing proteins, along with bead beating for mechanical disruption of the cells. However, the results obtained on three independent replicates were largely unsatisfactory (Figure 1, TUC–MD). A sonication step was therefore added to aid cell lysis, providing a slight improvement in extraction yield (three-fold higher than without sonication), but still far from being satisfactory (Figure 1, TUC–MD + S). In order to boost protein solubilization and improve disruption of the yeast cell wall, a stronger SDS-based extraction buffer was used, and sequential bead-beating steps were alternated with either freezing or boiling. This harsher procedure led to a dramatic increase in protein extraction yield (>400-fold higher compared to TUC–MD, according to relative abundance estimation based on densitometric values; Figure 1, SDS–TS + MD).

We next assessed the suitability of the obtained protein extracts for 2D-PAGE analysis of the *R. mucilaginosa* whole proteome. Given the low compatibility of SDS with IEF separation, the protein extracts were diluted 1:10 in TUC buffer to bring SDS concentration below 0.1%. Furthermore, the performances of medium–small-format gels were evaluated. As a result (Figure 2, top), 2D map profiles revealed focusing problems in the acidic zone, along with a large and disturbing interference in the bottom-left edge of the gel, possibly due to binding of fluorescent dyes to some contaminating molecules. To overcome this issue, the protein extract was cleaned up using an established commercial method to eliminate interfering substances, and large-format gels were employed to increase 2D-PAGE resolution. This enabled us to obtain patterns of considerably higher quality and complexity, and underlined the need for protein clean-up and large-format gels to generate satisfactory 2D-PAGE maps of the *R. mucilaginosa* whole proteome (Figure 2, bottom).

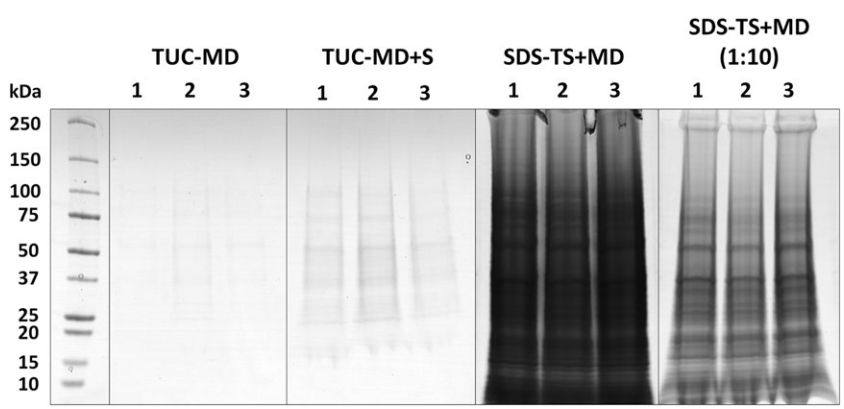

Figure 1. Comparison of SDS–PAGE patterns obtained with protein extraction methods of increasing harshness levels. Results from three independent *R. mucilaginosa* C2.5 t1 cultures are shown. For the SDS–TS + MD extraction method, both neat and diluted extracts are displayed

DOI: 10.1002/yea

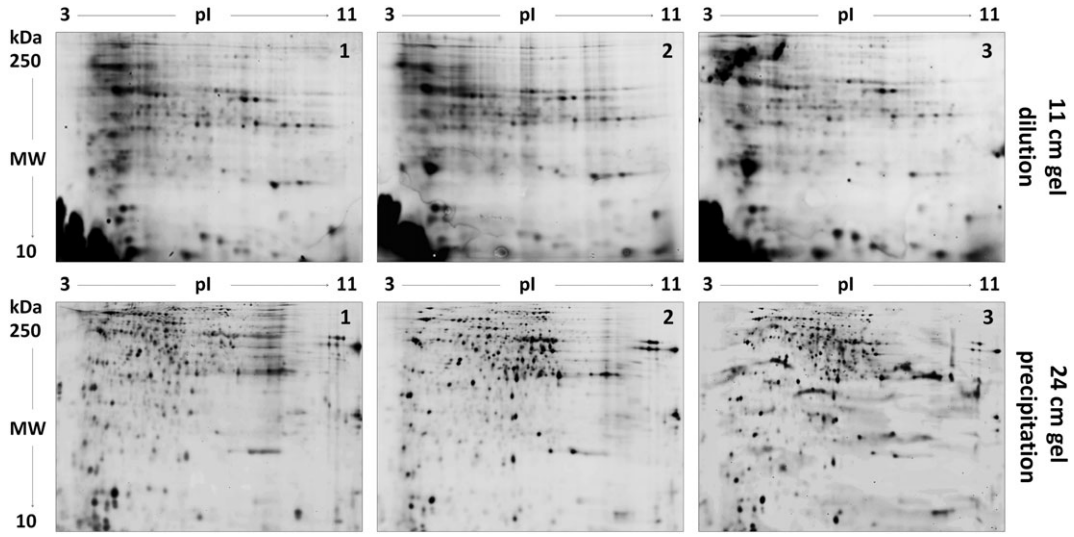

Figure 2. 2D-PAGE profiles of protein extracts from three independent *R. mucilaginosa* C2.5t1 cultures. Extracts were either diluted 10-fold and separated on small-format gels (top) or cleaned up, resuspended in TUC buffer and separated on large-format gels (bottom)

Application of 2D-DIGE to wild-type and mutant strains

As a further investigation, the ability of 2D-DIGE to identify proteins of interest in *R. mucilaginosa* was evaluated. The production of carotenoids is one of the characterizing features of the species. Therefore, the two pilot experiments described below were designed to assess the ability of this technique to uncover possible protein abundance differences in the carotenogenic pathway.

Pilot experiment A

Based on the kinetics of carotenoid production during growth in YEPGLY (Figure 3), we assessed the performance of 2D-DIGE when employed to detect protein expression changes at the different growth stages of interest (16, 40 and 72 h) in the parental strain C2.5 t1. The rationale behind this experimental design was that, if the enzymes involved in the biosynthesis of carotenoids change their expression level during growth and

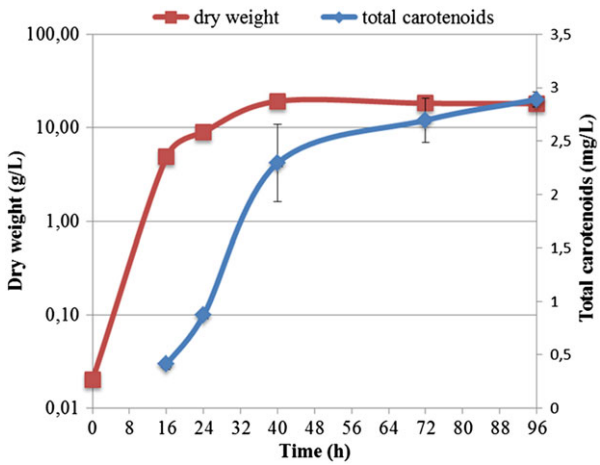

Figure 3. Growth and carotenoid production. Growth was measured by evaluating the dry weight of biomass (shown on a log scale). Total carotenoids are expressed as mg/L β-carotene equivalents. Data are mean ± SD of six independent experiments; where not visible, bars lie within the symbols

DOI: 10.1002/yea

carotenoid accumulation, they could be highlighted among differentially expressed proteins at the different sampling times. Overlay images of the 2D-DIGE comparisons are provided in Figure 4 (top). Upon DeCyder analysis, only 22 differential spots showed a consistently higher intensity in one of the three time points analysed, when compared to at least one of the other time points. Among them, six spots, which could be evidently tracked in the preparative gels, were selected for cutting and MS analysis (circled in Figure 4, top). The MS results are described in the following section.

Pilot experiment B

The second pilot experiment was aimed at comparing the protein expression profile of the parental strain C2.5t1 with those of two mutants, 400A15 and 200A6, that differ from the parental strain in the amount and type of carotenoids produced. Strain 400A15 over-produces β-carotene (Cutzu *et al.*, 2013a, 2013b), while 200A6 is unable to produce detectable amounts of the main carotenoids found in the parental strain (data not shown). Based on the assumption that the phenotypic characteristics of the mutants could be originated by alterations in different stages of the biosynthetic pathway, we employed these three strains to identify the enzymes involved in carotenoid production. To this aim, the three strains were analysed after 72 h of growth in YEPGLY medium. Representative overlay images of the 2D-DIGE comparisons are provided in Figure 4 (bottom). Upon DeCyder analysis, 119 differential spots showed consistently higher intensity in one of the three strains when compared to at least one of the other strains. Among them, 26 spots were selected for cutting and MS analysis (circled in Figure 4, bottom), based on two conditions: (a) to be clearly visible on the preparative gels; and (b) to be either more intense in C2.5t1 and 400A15 when compared to 200A6, or more intense in 400A15 when compared to both C2.5t1 and 200A6. The results of MS analysis are described in the following section.

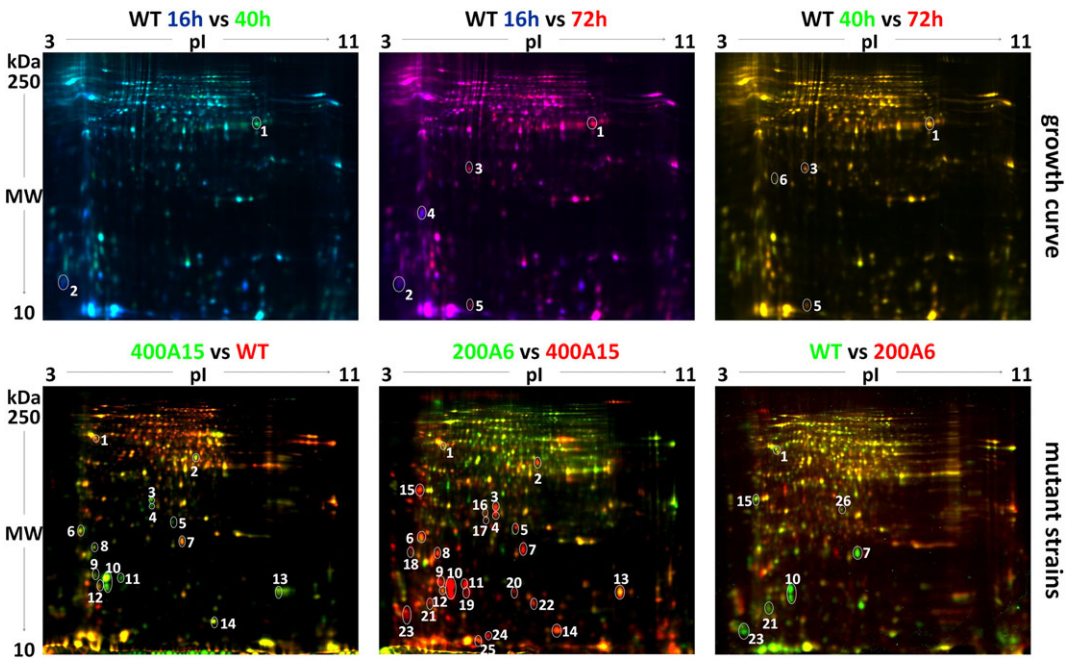

Figure 4. 2D-DIGE experiments. (Top) 2D-DIGE overlay images from the growth curve analysis experiment: dark blue, green and red characters indicate Cy2, Cy3 and Cy5 labelling, respectively; overlay patterns are marked in light blue (16 vs 40 h), lilac (16 vs 72 h) and yellow (40 vs 72 h); differential spots cut and analysed by LC–MS/MS are highlighted and numbered. Strain C2.5t1 is indicated as WT. (Bottom) 2D-DIGE representative overlay images from the mutant strains comparison experiment: green and red characters indicate Cy3 and Cy5 labelling, respectively; overlay patterns are marked in yellow; differential spots cut and analysed by LC–MS/MS are highlighted and numbered

DOI: 10.1002/yea

Comparison of database search strategies for mass spectrometry identification of spots

In view of the often limited availability of sequence information on non-conventional yeasts, another relevant aim of our work was to find the best database-searching strategy for improving the outcome of a gel-based proteomic study. Therefore, a critical comparison was carried out among three different sequence databases, as follows. The spectra obtained from MS analysis of the peptide mixtures deriving from the 32 differential spots described above (pilot experiments A and B) were searched against three different sequence databases: S-DB (UniProtKB sequences assigned to the order *Sporidiobolales*, to which *R. mucilaginosa* belongs); C-DB (UniProtKB protein sequences, without any taxonomic filter, mapping to the carotenoid biosynthesis pathway); and G-DB (sequences generated upon genome sequencing of *R. mucilaginosa* strain C2.5t1, recently deposited and published; Deligios *et al.*, 2015). As a result, G-DB clearly outperformed the competing databases, even though both S- and C-DB provided a few unique peptide identifications (Figure 5).

Then, the percentage of spots with a reliable protein identification (i.e. with at least two unique peptides detected) was investigated for each database search. G-DB allowed the achievement of reliable protein identifications for about half of the spots analysed, versus one-fifth and zero for S- and C-DB, respectively (Figure 6).

Based on these results, the availability of genomic sequences from the same strain analysed by proteomics was demonstrated to provide a significant improvement in protein identification performances, leading to 3.5-fold more peptide identifications when compared to the deposited sequences from taxonomically related yeasts, presumably due to significant changes in genome coding sequences.

On the whole, 3 and 15 protein spots returned a valid identification (FDR $>$ 1%, at least two unique peptides) for the two pilot experiments, A and B (described in section 3.2), respectively. However, even when the results obtained using G-DB and S-DB were merged, only in very few cases it was possible to assign a specific and unambiguous protein identity to a spot (e.g. spots 1, 6 and 19 in the 2D maps of pilot experiment B). Conversely, many different proteins with a few peptides each were detected in most spots, sometimes with the most abundant ones being uncharacterized proteins. Surprisingly, no carotenogenic enzymes could be identified, not even using C-DB. The complete data concerning spot identifications are given in Table S1 (see supporting information).

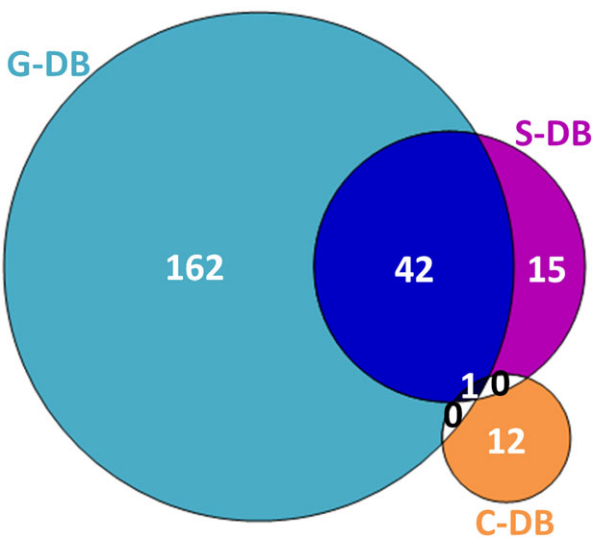

Figure 5. Distribution of all peptide identifications achieved from gel spots. Venn diagram illustrating the identifications achieved with the experimental genome (G-DB), UniProtKB Sporidiobolales (S-DB) or UniProtKB carotenogenesis-related proteins (C-DB) as sequence databases

DOI: 10.1002/yea

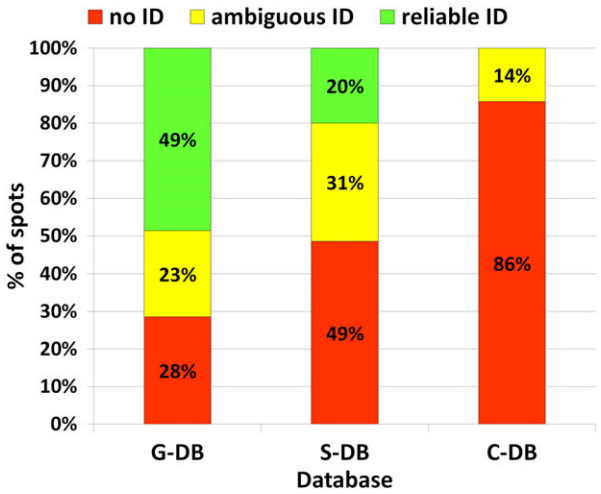

Figure 6. Reliability of spot identifications according to the different databases used. The percentages for which reliable (proteins with at least two unique peptides), ambiguous (proteins with only one unique peptide) or no protein identifications were obtained are reported according to the experimental genome (G-DB), UniProtKB Sporidiobolales (S-DB) or UniProtKB carotenogenesis-related proteins (C-DB) databases

Application of shotgun proteomic workflows

Three different combined proteomic workflows with varying degrees of sensitivity, affordability and technical complexity were evaluated for their performances in assessing the changes occurring during *R. mucilaginosa* growth. In order to maximize protein expression differences, we selected the two extreme time points of the growth curve, i.e. 16 and 72 h. Specifically, we compared a newer gel-free approach requiring high resolution mass spectrometry with a more labour-intensive and affordable GeLC–MS/MS approach. Furthermore, we decided to also analyse the strongly orange-coloured residual pellet to verify whether a relevant portion of the cell proteins (and/or, possibly, some specific protein classes) could not be properly partitioned and solubilized in the so-called 'protein extract' and remained 'trapped' in the pellet; to ensure an adequate analysis depth, a very high resolution mass spectrometer was employed in this latter case. Concerning analysis of MS data, a merged database was generated by appending sequences from S- and C-DB to those of G-DB, in order to maximize the search space. The three approaches were constructed as follows:

1 *EGIT*. The first strategy was named 'EGIT', as the protein extract was subjected to gel-based clean-up/fractionation and the peptide mixtures obtained from each gel slice by in-gel digestion were analysed using an XCT Ultra Ion Trap as the mass spectrometer.
2 *EFOV*. The second strategy was named 'EFOV', as the protein extract was cleaned up and digested according to the FASP approach, and the peptide mixture was analysed using an LTQ-Orbitrap Velos as the mass spectrometer.
3 *PGQE*. The third strategy was named 'PGQE', as the analysis was carried out on the residual cell pellet obtained after the last centrifugation step in the protein extraction phase, which was cleaned up and separated by short gel electrophoresis, and the peptide mixture obtained upon in-gel digestion was analysed using a Q-Exactive as the mass spectrometer.

At first, we assessed the results produced by each of the three strategies by merging 72 and 16 h data and comparatively evaluated as shown in Figure 7.

In general, the PGQE strategy enabled the identification of a higher number of proteins, corresponding in turn to a higher number of metabolic and functional features detected, followed by EFOV and EGIT. Globally speaking, the three strategies produced heterogeneous and complementary results, since only about one-quarter of the identified features were common to all approaches, with PGQE and EFOV providing the

DOI: 10.1002/yea

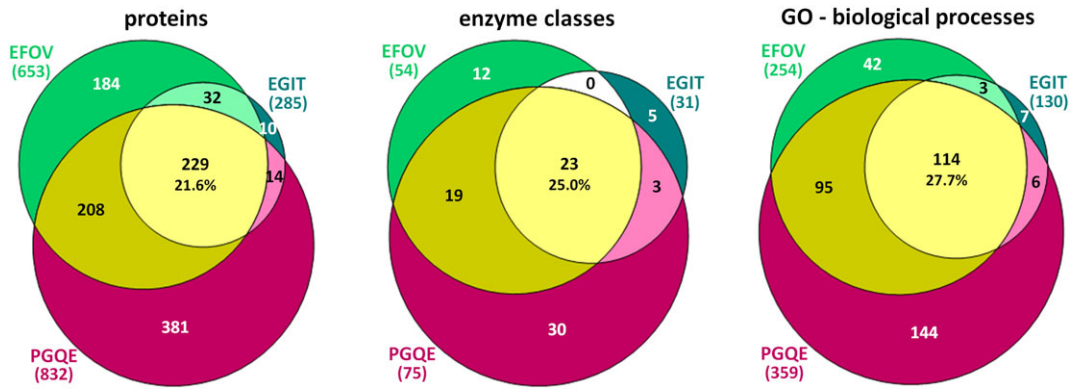

Figure 7. Distribution of proteins, enzyme classes and biological processes among the three non-2D gel-based strategies. Venn diagrams depicting distribution of proteins (left), enzyme classes (middle) and GO biological processes (right) among the three different shotgun proteomic strategies being compared; percentages of common features are indicated in the overlapping yellow area, while the total number of identified features for each individual approach is shown in parentheses

higher amounts of both total and unique identifications (Figure 7). In total, shotgun proteomic analysis led to the detection of $>12\%$ of the ORFs included in the G-DB.

To investigate on qualitative differences, we calculated the percentage of proteins containing transmembrane domains, in order to highlight a possible enrichment in membrane/hydrophobic proteins with one of the approaches. As a result, the amount of transmembrane proteins was quite similar for the three strategies, in the range 10–13% of the total (slightly higher for PGQE). Then, we assigned each protein to a KEGG orthology group (KOG) and KOG data were imported into iPATH to generate metabolic maps associated with each strategy. As illustrated in Figure 8, PGQE provided the deepest metabolic map coverage, followed by EFOV and EGIT.

We then carried out a specific comparison between 70 and 16 h data, in order to identify functional features exhibiting abundance changes along the cell growth and carotenoid accumulation. Figure 9 reports KOGs (A) and GO-biological processes (B) consistently showing a differential expression between the two time points, according to the results achieved with the two best-performing techniques [EFOV and PGQE; further details are given in the legend to Figure 9, while the overall data are provided in Table S2 (see supporting information)]. Specifically, 19 KOGs and 18 biological processes increased from 16 to 72 h of growth, whereas 46 KOGs and 42 biological processes decreased. Again, no protein functions strictly belonging to the carotenoid biosynthesis pathway were detected, either in general or as differentially expressed.

Discussion

The proteomic study of non-conventional yeasts poses several problems relating to both cell structure, with a thick and complex cell wall, and the scarcity of data concerning genome sequencing and characterization. Nevertheless, the application of proteomics to these organisms may unveil features of significant interest for either environmental, biotechnological or health implications. Here, we assessed the performance of different proteomic approaches in characterizing the protein repertoire expressed by the non-conventional yeast *R. mucilaginosa*. Specifically, adding to traditional gel-based proteomics, three different proteomic worflows were also implemented and applied. Due to the combination of the two different analytical strategies, a fair coverage of the proteomic repertoire was expected. In fact, although gel-based proteomics has its advantages, including the ability to reveal some post-translationally modified proteins entailing charge or size changes (Westermeier *et al.*, 2008), shotgun proteomics usually exhibits higher sensitivity, with the key added ability to provide information on the whole proteomic profile (Otto *et al.*, 2014).

DOI: 10.1002/yea

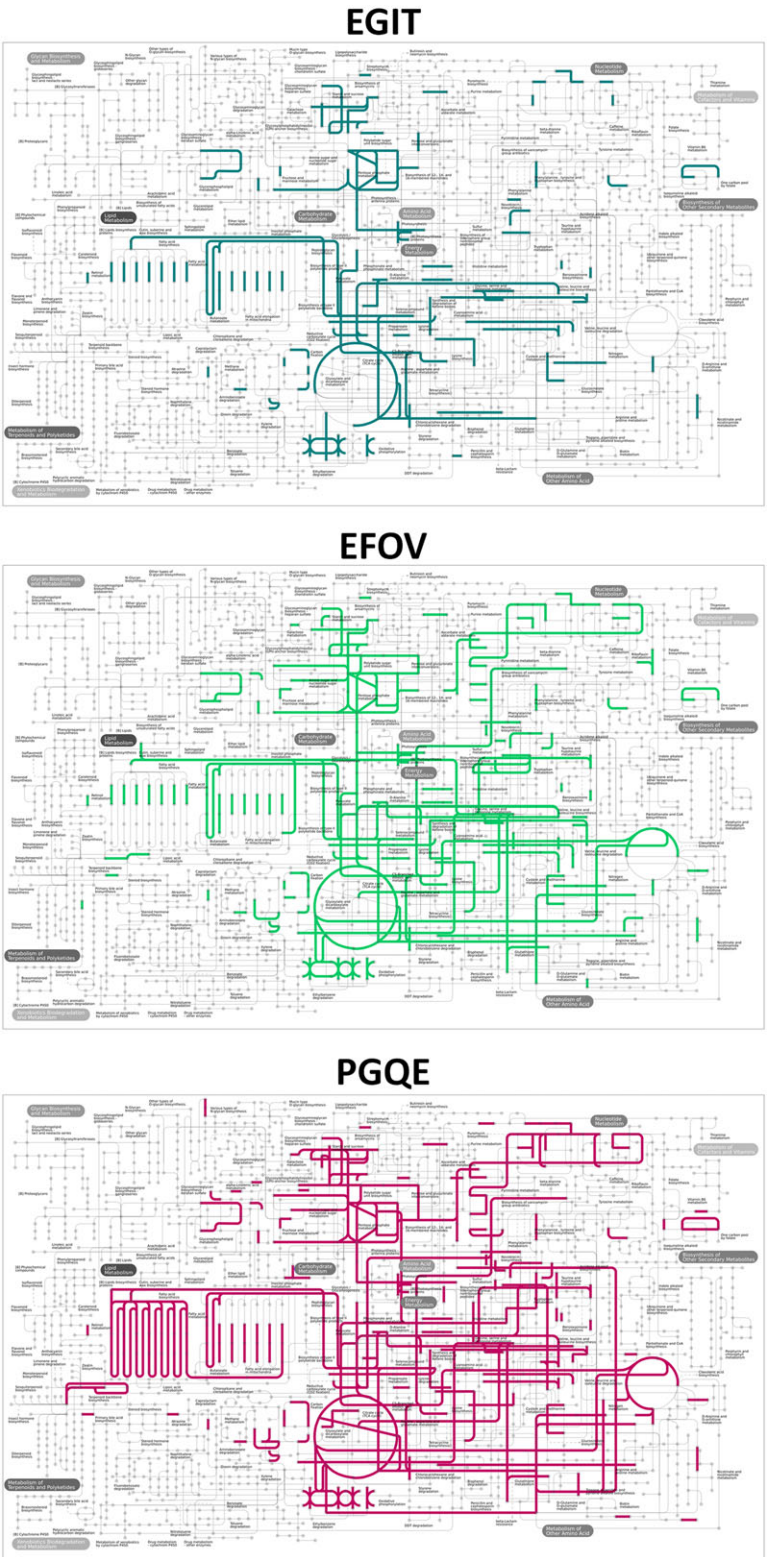

Figure 8. iPath metabolic pathways mapping all identified proteins: bars representing pathway steps are marked in blue-green (top), green (middle) and claret (bottom) for EGIT, EFOV and PGQE, respectively

DOI: 10.1002/yea

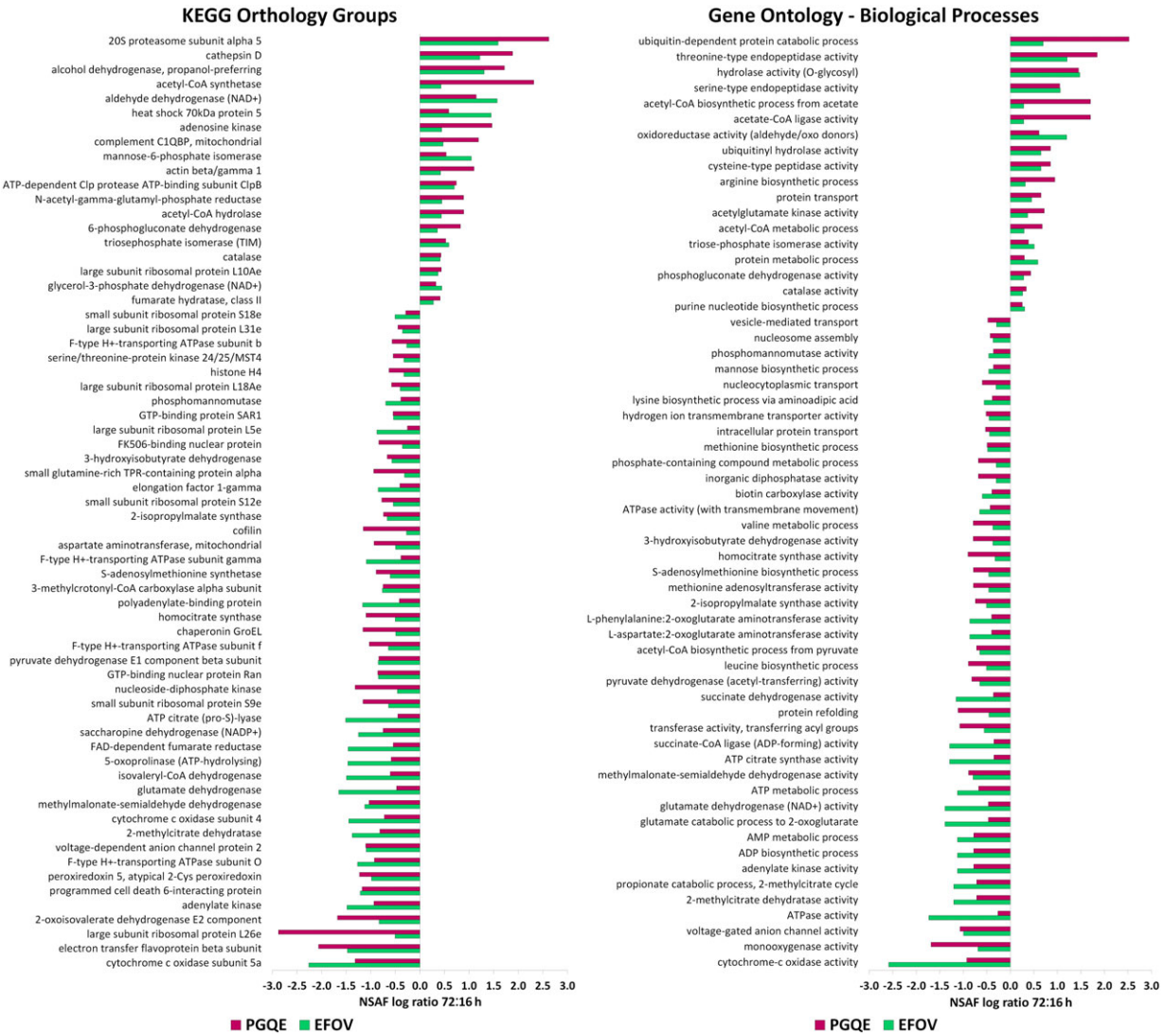

Figure 9. Bar graphs illustrating KEGG Orthology Groups (left) and Gene Ontology biological processes (right). The category distributions consistently show a differential expression between two growth curve time points (16 and 72 h) using the EFOV and PGQE methods. Quantitative comparisons were carried out by calculating the 72:16 h NSAF log ratio for each functional category, and features with log ratio > 0.25 or < -0.25 with both methods (as well as with a minimum of two peptides in at least one time point) are reported; data are ordered according to decreasing mean NSAF log ratio values

The first and significant problem encountered in proteomic analysis of *R. mucilaginosa* was the difficulty of obtaining an efficient cell lysis with a satisfactory protein extraction and solubilization efficiency. In fact, strong detergents and harsh physical treatments were required to break the cell walls, and a protein precipitation step was needed to remove interfering substances in order to obtain acceptable electrophoretic patterns. This notwithstanding, we observed that the insoluble, pelleted cell debris remained pigmented, while the extract was only lightly coloured. There is therefore the possibility that some interesting proteins, including those related to carotenoid biosynthesis or to the biosynthesis of other relevant proteins, may remain 'trapped' in the pellet and escape extraction and analysis (Guo *et al.*, 2014). In fact, when this fraction was specifically investigated (by PGQE), a slight enrichment in hydrophobic proteins was observed. Nevertheless, no carotenogenic enzymes were identified, even in this fraction, leaving open the possibility that this

DOI: 10.1002/yea

function may be tightly associated with some insoluble component.

Concerning 2D-DIGE, the application of this technique to cleaned protein extracts did successfully highlight significant spot differences among patterns, in the case of both different strains and different growth stages of the same strain. Nevertheless, when cut and subjected to MS analysis, many spots returned identifications related to abundant proteins or to uncharacterized proteins. This problem is likely caused by the poor availability of a well-annotated database, which hampers the successful identification of less common or characterized proteins. In addition, the protein precipitation step, which was found to be necessary to minimize interference on gel-based separation and detection, may have led to the selective loss of some protein categories (Tanca *et al*., 2013a). In virtue of their wider proteome coverage, an advantage of shotgun proteomics when compared to 2D gel-based workflows lies in the ability to provide indirect information on a pathway of interest by means of the 'context' proteins, i.e. proteins belonging to interacting or biochemically related pathways (Otto *et al*., 2014). An example of this is provided by the pathway maps illustrated in Figure 7.

Two of the proteomic analysis workflows assessed here (EGIT and PGQE) integrate a preliminary gel-based protein separation, facilitating elimination of most of the interfering molecules, which remain trapped into the gel matrix. In addition, these have the added advantage of providing sample prefractionation. These two aspects may be especially relevant when dealing with a yeast such as *Rhodotorula* spp., when considering that, due to the cell characteristics, some contaminants are present, such as lipids, complex polysaccharides and pigments, that may be extracted together with the proteins and act as interfering contaminants when analysing the protein mixture. Methods that can enable the removal of such substances, such as those entailing a gel-based separation, although performing less well in terms of identified proteins, may be preferable due to these 'sample-cleaning' features. Nevertheless, the EGIT approach (elsewhere also named GeLC–MS/MS), although being the most labour-intensive and time-consuming in terms of operator hands-on time, provided less satisfactory results in terms of protein identifications. In this specific case, however, it should be considered that a low resolution and low sensitivity mass spectrometer was used. We therefore cannot rule out that this sample preparation approach may provide better results when combined with a higher performance mass spectrometer.

On the other hand, EFOV, based on the increasingly used FASP procedure, has the advantage of reproducibility and scalability, and can generate a richer protein identification dataset. Nevertheless, it may cause technical drawbacks and performance problems to the LC–MS/MS equipment or separation accessories, due to the persistence of 'impurities' or contaminating molecules, which may cause build-ups and alter or impair mass spectra quality (data not shown). In fact, EFOV does not include sample prefractionation steps, although the use of a filter device for performing sample digestion can somewhat enable the removal of small contaminant molecules.

Adding to sample preparation, another crucial aspect in generating reliable proteomic data for non-model organisms (as *R. mucilaginosa*) is the choice of the proper sequence database to be used for protein identification (Armengaud *et al*., 2014). The data presented here further highlight how significant protein identification issues can arise when dealing with poorly characterized microorganisms, due to difficulties in matching experimental mass spectra with the deposited sequences belonging to related, better-characterized species/strains. Accordingly, when using a matched database produced upon genome sequencing of the specific microbial strain under study, a dramatic increase in the number of identified proteins can be obtained, as demonstrated here when interrogating the G-DB. This is in line with previous results from other *Rhodotorula* strains (Tanca *et al*., 2013b), as well as from other yeasts, such as *Rhodosporidium toruloides*. In this latter case, an LC–MS/MS dataset was initially searched against a 'generic' yeast dataset, with no more than 184 proteins identified (Liu *et al*., 2009), while a subsequent re-analysis of the same dataset with the translated genome obtained by next-generation DNA sequencing reached >3100 protein identifications, corresponding to a dramatic 17-fold increase (Zhu *et al*., 2012). Therefore, the application to non-model yeasts of a 'proteogenomic' workflow, i.e. one in which genome sequencing data are used to improve proteomic results and, in

DOI: 10.1002/yea

turn, proteomic data are employed to validate the actual expression of sequenced genes, can clearly provide significant improvements. On the other hand, this is counterbalanced by the fact that whole-genome sequencing of an eukaryotic microorganism implicates significantly higher costs, due to genome size. Also, and equally important, more serious bioinformatic issues arise in this case, e.g. concerning read assembly, ORF finding and functional annotation, to name a few, when compared to the simpler prokaryotes (Armengaud *et al*., 2014). As a confirmation of these issues, most of the proteins found in this study (especially those detected as differentially expressed with gel-based and/or gel-free approaches) could not be reliably associated with a specific function. The reason for this can possibly be related to a poor alignment with the currently deposited protein sequences (either by the very low homology with known sequences, or due to the presence of many truncated genes in the genome draft), as well as to the absence of a functional annotation in the homologous proteins from related yeast species (e.g. *R. toruloides*). The presence of truncated gene sequences may also partially explain the identification of peptides from different proteins within the same spot, as observed in the 2D-DIGE analysis.

It was, however, surprising that we did not identify any of the enzymes known as being involved in the carotenogenic pathway, one of the most interesting biotechnological features of this yeast, in spite of the use of many different molecular approaches evaluated here, including the shotgun proteomics and genome sequencing and annotation (with four sequenced genes likely matching with specific carotenogenic enzymes, according to BLASTp alignment). Concerning the reasons that may lie behind this result, it is interesting to highlight the very low intracellular concentration reported for carotenogenic enzymes (Sandmann, 1997), which might have prevented their identification with the different proteomic approaches employed here. In addition, the heterogeneity of *R. mucilaginosa* protein sequences compared to those of other yeasts might also have accounted for this result, worsening the chances of their detection and identification.

This notwithstanding, the combined proteomic workflows applied here provided valuable information for the generation of a proteome database that may assist further studies on *R. mucilaginosa*. Protein mapping to metabolic pathways provided a good coverage of central metabolism, and the implementation of PGQE strategy, which enabled a slight enrichment in membrane proteins, highlighted a segment of the terpenoid backbone pathway (Figure 8, bottom). The identification of functional features showing consistent abundance changes throughout cell growth returned useful information on the regulation of different metabolic pathways at two different growth stages. In particular, as expected during exponential growth on glycerol-containing medium, pyruvate dehydrogenase and proteins involved in acetyl-CoA biosynthesis from pyruvate were upregulated in cells sampled after 16 h of growth. In contrast, cells sampled at 72 h of growth showed an increase in the abundance level of proteins involved in cytosolic acetyl CoA production from acetate (aldehyde dehydrogenase NAD^+, acetate–CoA ligase activity, acetyl-CoA biosynthetic process from acetate and acetyl-CoA metabolic processes). Cytosolic acetyl-CoA is the source for fatty acids and sterol but also for carotenoid biosynthesis (Chen *et al*., 2012). Thus, the increase in the expression level of enzymes involved in the biosynthesis of cytosolic acetyl-CoA is in accordance with the accumulation of carotenoids in stationary phase. Moreover, in accordance with the upregulation of enzymes involved in the stress response during carotenogenesis (Barbachano-Torres *et al*., 2014; Martinez-Moya *et al*., 2015), at the stationary phase of growth there was a slight but consistent increase in the expression levels of catalase and heat shock protein HSP70. According to Martinez-Moya *et al*. (2015), the higher abundance of enzymes involved in the response to stress in stationary phase would be related to the induction of carotenogenesis.

In conclusion, proteomics is able to provide a wealth of information on numerous protein functions and biosynthetic pathways of *R. mucilaginosa*, as demonstrated by the vast dataset generated and by the information obtained on different biosynthetic pathways, and on how these change upon growth or upon the mutation of phenotypic traits. Nevertheless, when analysing the data for one of its most biotechnologically relevant pathways, i.e. carotenoid production, the level of information gathered with the different technical approaches did not provide satisfactory information, due to the low expression level of

DOI: 10.1002/yea

carotenogenic enzymes but also to poor alignments of the MS spectra with the currently deposited protein sequences. Therefore, dedicated efforts in characterization of the species at the genome level, together with a careful annotation of genome sequences, might be required to improve research efforts aimed to exploit the biotechnological potential offered by non-conventional yeasts.

Acknowledgements

This study was supported by Regione Autonoma della Sardegna (Grant No. LR7/07-2010, to I.M.).

References

Addis MF, Pisanu S, Preziosa E, *et al.* 2012. 2D DIGE/MS to investigate the impact of slaughtering techniques on postmortem integrity of fish filet proteins. *J Proteom* **75**: 3654–3664.

Aksu Z, Eren AT. 2005. Carotenoids production by the yeast *Rhodotorula mucilaginosa*: use of agricultural wastes as a carbon source. *Process Biochem* **40**: 2985–2991.

Armengaud J, Trapp J, Pible O, *et al.* 2014. Non-model organisms, a species endangered by proteogenomics. *J Proteom* **105**: 5–18.

Barbachano-Torres A, Castelblanco-Matiz LM, Ramos-Valdivia AC, *et al.* 2014. Analysis of proteomic changes in colored mutants of *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). *Arch Microbiol* **196**: 411–421.

Canli O, Erdal S, Taskin M, Kurbanoglu EB. 2011. Effects of extremely low magnetic field on the production of invertase by *Rhodotorula glutinis*. *Toxicol Ind Health* **27**: 35–39.

Cheirsilp B, Suwannarat W, Niyomdecha R. 2011. Mixed culture of oleaginous yeast *Rhodotorula glutinis* and microalga *Chlorella vulgaris* for lipid production from industrial wastes and its use as biodiesel feedstock. *Nat Biotechnol* **28**: 362–368.

Chen Y, Siewers V, Nielsen J. 2012. Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in *Saccharomyces cerevisiae*. *PLoS One* **7**: e42475.

Cutzu R, Clemente A, Reis A, *et al.* 2013a. Assessment of β-carotene content, cell physiology and morphology of the yellow yeast *Rhodotorula glutinis* mutant 400A15 using flow cytometry. *J Ind Microbiol Biotechnol* **40**: 865–875.

Cutzu R, Coi A, Rosso F, *et al.* 2013b. From crude glycerol to carotenoids by using a *Rhodotorula glutinis* mutant. *World J Microbiol Biotechnol* **29**: 1009–1017.

Deligios M, Fraumene C, Abbondio M, *et al.* 2015. Draft genome sequence of *Rhodotorula mucilaginosa*, an emergent opportunistic pathogen. *Genome Announc* **3**: e00201–15.

Fell JW, Stratzell-Tallman A. 1998. *Rhodotorula* F. C. Harrison. In *The Yeasts*, *A Taxonomic Study*, Kurtzman CP, Fell JW (eds), 4th edn. Elsevier: Amsterdam.

Galafassi S, Cucchetti D, Pizza F, *et al.* 2012. Lipid production for second generation biodiesel by the oleaginous yeast *Rhodotorula graminis*. *Bioresour Technol* **111**: 398–403.

Ghisaura S, Anedda R, Pagnozzi D, *et al.* 2014. Impact of three commercial feed formulations on farmed gilthead sea bream (*Sparus aurata* L.) metabolism as inferred from liver and blood serum proteomics. *Proteome Sci* **12**: 44.

Guo W, Tang H, Zhang L. 2014. Lycopene cyclase and phytoene synthase activities in the marine yeast *Rhodosporidium diobovatum* are encoded by a single gene, *crtYB*. *J Basic Microbiol* **54**: 1053–1061.

Hernández-Almanza A, Cesar Montanez J, Aguilar-González MA, *et al.* 2014. *Rhodotorula glutinis* as source of pigments and metabolites for food industry. *Food Biosci* **5**: 64–72.

Ignatova LV, Brazhnikova YV, Berzhanova RZ, Mukasheva TD. 2015. Plant growth-promoting and antifungal activity of yeasts from dark chestnut soil. *Microbiol Res* **175**: 78–83.

Irazusta V, Estévez C, Amoroso MJ, de Figueroa LIC. 2012. Proteomic study of the yeast *Rhodotorula mucilaginosa* RCL-11 under copper stress. *Biometals* **25**: 517–527.

Jarboui R, Baati H, Fetoui F, *et al.* 2012. Yeast performance in wastewater treatment: case study of *Rhodotorula mucilaginosa*. *Environ Technol* **33**: 951–960.

Johnson EA. 2013. Biotechnology of non-*Saccharomyces* yeasts – the Basidiomycetes. *Appl Microbiol Biotechnol* **97**: 7563–7577.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567–580.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Lahav R, Nejidat A, Abeliovich A. 2004. Alterations in protein synthesis and levels of heat shock 70 proteins in response to salt stress of the halotolerant yeast *Rhodotorula mucilaginosa*. *Antonie Van Leeuwenhoek* **85**: 259–269.

Li M, Liu GL, Chi Z, Chi ZM. 2010. Single cell oil production from hydrolysate of cassava starch by marine-derived yeast *Rhodotorula mucilaginosa* TJY15a. *Biomass Bioenergy* **34**: 101–107.

Li R, Zhang H, Liu W, Zheng X. 2011. Biocontrol of postharvest gray and blue mold decay of apples with *Rhodotorula mucilaginosa* and possible mechanisms of action. *Int J Food Microbiol* **146**: 151–156.

Libkind D, Gadanho M, van Broock M, Sampaio JP. 2008. Studies on the heterogeneity of the carotenogenic yeast *Rhodotorula mucilaginosa* from Patagonia, Argentina. *J Basic Microbiol* **48**: 93–98.

Liu H, Zhao X, Wang F, *et al.* 2009. Comparative proteomic analysis of *Rhodosporidium toruloides* during lipid accumulation. *Yeast* **26**: 553–566.

Maldonade IR, Rodriguez-Amaya DB, Scamparini ARP. 2012. Statistical optimization of cell growth and carotenoid production by *Rhodotorula mucilaginosa*. *Braz J Microbiol* **43**: 109–115.

Mannazzu I, Landolfo S, da Silva TL, Buzzini P. 2015. Red yeasts and carotenoid production: outlining a future for non-conventional yeasts of biotechnological interest. *World J Microbiol Biotechnol* **31**: 1665–1673.

Martinez-Moya P, Niehaus K, Alcaíno J, *et al.* 2015. Proteomic and metabolomic analysis of the carotenogenic yeast *Xanthophyllomyces dendrorhous* using different carbon sources. *BMC Genom* **16**: 289.

Moriya Y, Itoh M, Okuda S, *et al.* 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**: W182–185.

Otto A, Becher D, Schmidt F. 2014. Quantitative proteomics in the field of microbiology. *Proteomics* **14**: 547–565.

Paulo JA, Kadiyala V, Brizard S, *et al.* 2013. Short gel, long gradient liquid chromatography tandem mass spectrometry to discover urinary biomarkers of chronic pancreatitis. *Open Proteom J* **6**: 1–13.
Pisanu S, Marogna G, Pagnozzi D, *et al.* 2013. Characterization of size and composition of milk fat globules from Sarda and Saanen dairy goats. *Small Rumin Res* **109**: 141–151.
Rajpert L, Skłodowska A, Matlakowska R. 2013. Biotransformation of copper from Kupferschiefer black shale (Fore-Sudetic Monocline, Poland) by yeast *Rhodotorula mucilaginosa* LM9. *Chemosphere* **91**: 1257–1265.
Sandmann G. 1997. High level expression of carotenogenic genes for enzyme purification and biochemical characterization. *Pure Appl Chem* **69**: 2163–2168.
Tampitak S, Louhasakul Y, Cheirsilp B, Prasertsan P. 2015. Lipid production from hemicellulose and holocellulose hydrolysate of palm empty fruit bunches by newly isolated oleaginous yeasts. *Appl Biochem Biotechnol* **176**: 1801–1814.
Tanca A, Pagnozzi D, Falchi G, *et al.* 2011. Application of 2D DIGE to formalin-fixed, paraffin-embedded tissues. *Proteomics* **11**: 1005–1011.
Tanca A, Pagnozzi D, Burrai G, Polinas M. 2012. Comparability of differential proteomics data generated from paired archival fresh-frozen and formalin-fixed samples by GeLC–MS/MS and spectral counting. *J Proteom* **77**: 561–576.
Tanca A, Biosa G, Pagnozzi D, *et al.* 2013a. Comparison of detergent-based sample preparation workflows for LTQ–Orbitrap analysis of the *Escherichia coli* proteome. *Proteomics* **13**: 2597–2607.
Tanca A, Palomba A, Deligios M, *et al.* 2013b. Evaluating the impact of different sequence databases on metaproteome analysis: insights from a lab-assembled microbial mixture Martens L. (ed.) *PLoS One* **8**: e82981.
Tanca A, Pisanu S, Biosa G, *et al.* 2013c. Application of 2D DIGE to formalin-fixed diseased tissue samples from hospital repositories: results from four case studies. *Proteom Clin Appl* **7**: 252–263.
Tanca A, Abbondio M, Pisanu S, *et al.* 2014a. Critical comparison of sample preparation strategies for shotgun proteomic analysis of formalin-fixed, paraffin-embedded samples: insights from liver tissue. *Clin Proteom* **11**: 28.
Tanca A, Palomba A, Pisanu S, *et al.* 2014b. A straightforward and efficient analytical pipeline for metaproteome characterization. *Microbiome* **2**: 1–16.
Taskin M. 2013. Co-production of tannase and pectinase by free and immobilized cells of the yeast *Rhodotorula glutinis* MP-10 isolated from tannin-rich persimmon (*Diospyros kaki* L.) fruits. *Bioprocess Biosyst Eng* **36**: 165–172.
Westermeier R. 2006. Sensitive, quantitative, and fast modifications for Coomassie blue staining of polyacrylamide gels. *Proteomics* **6** (suppl 2): 61–64.
Westermeier R, Naven T, Höpker HR. 2008. Proteomics in Practice: A Guide to Successful Experimental Design. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
Wirth F, Goldani LZ. 2012. Epidemiology of *Rhodotorula*: an emerging pathogen. *Interdiscip Perspect Infect Dis* **2012**: 465717.
Wiśniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. *Nat Methods* **6**: 359–362.
Yamada T, Letunic I, Okuda S, *et al.* 2011. iPath2.0: interactive pathway explorer. *Nucleic Acids Res* **39**: W412–415.
Zhang H, Ge L, Chen K, *et al.* 2014. Enhanced biocontrol activity of *Rhodotorula mucilaginosa* cultured in media containing chitosan against postharvest diseases in strawberries: possible mechanisms underlying the effect. *J Agric Food Chem* **62**: 4214–4224.
Zhang H, Wang L, Dong Y, *et al.* 2008. Control of postharvest pear diseases using *Rhodotorula glutinis* and its effects on postharvest quality parameters. *Int J Food Microbiol* **126**: 167–171.
Zhu Z, Zhang S, Liu H, *et al.* 2012. A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. *Nat Commun* **3**: 1112.
Zybailov B, Mosley AL, Sardiu ME, *et al.* 2006. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* **5**: 2339–2347.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Table S1. Protein identifications from 2-D DIGE spots using G-DB and S-DB as sequence databases, respectively.

Table S2. Complete data on protein identifications from the EGIT, EFOV and PGQE strategies.

DOI: 10.1002/yea