Yeast
Yeast 2007; **24**: 667–679.
Published online 29 May 2007 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/yea.1499

Research Article

Metabolic footprinting as a tool for discriminating between brewing yeasts

Georgina A. Pope[1]*, Donald A. MacKenzie[1], Marianne Defernez[2], Miguel A. M. M. Aroso[1], Linda J. Fuller[1], Fred A. Mellon[2], Warwick B. Dunn[3], Marie Brown[3], Royston Goodacre[3], Douglas B. Kell[3], Marcus E. Marvin[4], Edward J Louis[4] and Ian N. Roberts[1]

[1] *National Collection of Yeast Cultures, Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK*
[2] *Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK*
[3] *School of Chemistry, Manchester Interdisciplinary Biocentre, 131 Princess Street, University of Manchester, Manchester M1 7DN, UK*
[4] *Institute of Genetics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK*

*Correspondence to:
Georgina A. Pope, National Collection of Yeast Cultures, Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK.
E-mail:
georgina.pope@bbsrc.ac.uk

Received: 12 January 2007
Accepted: 24 March 2007

Abstract

The characterization of industrial yeast strains by examining their metabolic footprints (exometabolomes) was investigated and compared to genome-based discriminatory methods. A group of nine industrial brewing yeasts was studied by comparing their metabolic footprints, genetic fingerprints and comparative genomic hybridization profiles. Metabolic footprinting was carried out by both direct injection mass spectrometry (DIMS) and gas chromatography time-of-flight mass spectrometry (GC–TOF–MS), with data analysed by principal components analysis (PCA) and canonical variates analysis (CVA). The genomic profiles of the nine yeasts were compared by PCR–restriction fragment length polymorphism (PCR–RFLP) analysis, genetic fingerprinting using amplified fragment length polymorphism (AFLP) analysis and microarray comparative genome hybridizations (CGH). Metabolomic and genomic analysis comparison of the nine brewing yeasts identified metabolomics as a powerful tool in separating genotypically and phenotypically similar strains. For some strains discrimination not achieved genomically was observed metabolomically.

Keywords: ***Saccharomyces*; metabolic footprinting; metabolomics; direct injection mass spectrometry; GC–TOF–MS; AFLP**

Supplementary material for this article can be found at: **http://www.interscience.wiley.com/jpages/0749-503X/suppmat/**

Introduction

Within the industrial sector there is a continuing need for new methods for the improved characterization of yeast strains. For example, it is important to be able to separate brewing strains from non-brewing 'wild' yeast strains and also to discriminate between closely-related brewing strain types. Traditional characterization methods, based on biochemical, morphological and physiological criteria, deduced that top fermenting ale strains belong to *Saccharomyces cerevisiae*, while their bottom fermenting lager counterparts are grouped in the species *S. pastorianus* (Deak and Beuchat, 1996). Genome sequencing has determined that *S. pastorianus* strains are hybrids of *S. cerevisiae* and either *S. bayanus* or *S. monacensis*, an old species grouped now in *S. pastorianus* (Hansen and Kiellandbrandt, 1994; Vaughan-Martini and Kurtzman, 1985). The complexity grew when it was discovered that *S. bayanus* (CBS 380) is a hybrid of *S. cerevisiae* and

S. uvarum (Nguyen *et al*., 2000). Rainieri *et al*. (2006) deduced that there are pure strains of *S. bayanus* (e.g. NBRC 1948) and *S. uvarum* (CBS 7001) as well as hybrid lines that contain a third, as-yet unidentified lager genome; hybrid lines identified included *S. cerevisiae*/*S. bayanus*/Lager, *S. bayanus*/*S. uvarum*/Lager and *S. cerevisiae*/*S. bayanus*/*S. uvarum*/Lager. Rainieri *et al*. (2006) suggested that the *S. pastorianus* name be used for multiple genetic lines that contain the *S. cerevisiae* genome.

The identification and classification of yeast isolates is a continuous requirement and is normally applied at the genomic level. In the recent past, various molecular techniques have been used to separate, classify and identify brewing yeasts; including the interlinked species *S. cerevisiae* and *S. pastorianus* (Tornai-Lehoczki and Dlauchy, 2000; Wightman *et al*., 1996; Yamagishi *et al*., 1999). However, the genetic complexity inherent in the brewing yeast species can result in misclassification and genetic methods do not always allow reliable separation of closely related strains. The metabolome (collection of low molecular weight organic and inorganic chemical species present in a cell or biological system) can detect genotypic and phenotypic differences in yeast (Oliver, 1998; Raamsdonk *et al*., 2001; Himmelreich *et al*., 2003, 2005) and is expected theoretically, and has been shown experimentally, to have a greater discriminatory power than transcriptomics and proteomics (Kell and Westerhoff, 1986; Urbanczyk-Wochniak *et al*., 2003). The aim of the present work was thus to use a metabolomics approach for the chemotaxonomy of brewing yeasts.

The metabolome is considered to be a more functional level at which to study biological systems, as it represents the final downstream product of gene expression (Dunn *et al*., 2005; Goodacre *et al*., 2004; Kell, 2004, 2006; Maharjan and Ferenci, 2005). Protocols involving rapid quenching and extraction of intra-cellular metabolites (the endometabolome) into organic solvents are employed to study intracellular metabolism (Maharjan and Ferenci, 2003). However, this is technically and economically demanding (Villas-Bôas *et al*., 2005) because of rapid metabolic fluxes and the complexity of the metabolome (size, physical/chemical properties, large concentration ranges). Metabolic footprinting (Allen *et al*., 2003, 2004; Kell *et al*., 2005) provides an alternative that does not rely on the measurement of intracellular metabolites but on the monitoring of the exometabolome, a combination of the metabolites secreted from the intracellular volume and any unused growth medium components, and as such can be performed without the requirement for quenching and extraction. A metabolite-rich medium is employed to 'probe' intracellular metabolism, with unused medium components and excreted metabolites being detected in the metabolic footprint. Metabolic footprinting has previously proved useful in classifying gene functions by comparing the collection of single knockout strains of *S. cerevisiae* (Allen *et al*., 2003), in mode of action studies (Allen *et al*., 2004), in studying the *Saccharomyces* population status of wine fermentations (Howell *et al*., 2006) and in accessing bacterial fibre degradation (Villas-Bôas *et al*., 2006). Secretion of metabolites can be optimized by supplementing the medium with a metabolite cocktail to stimulate 'overflow metabolism' (Allen *et al*., 2003; Kell *et al*., 2005). A number of analytical technologies can be employed to study metabolic footprints (Dunn *et al*., 2005). Direct injection electrospray mass spectrometry (DIMS) and gas chromatography-time of flight-mass spectrometry (GC–TOF–MS) were chosen to be used in this study to assess whether different brewing yeast strains could be discriminated from each other on the basis of their metabolic footprints.

The genetic architecture of the yeasts was examined by AFLP, whole genome microarray analysis and PCR–restriction fragment length polymorphism (PCR–RFLP) analysis. AFLP is a PCR-based genotyping method (patented by Keygene, The Netherlands) that has been developed for differentiating at both the inter- and intra-specific levels (Koeleman *et al*., 1998). AFLP patterns are a representation of the whole genome and have been used for strain typing and species identification purposes for a variety of microorganisms, including yeasts of medical and industrial importance (Azumi and Goto-Yamamoto, 2001; de Barros Lopes *et al*., 2002). In this study, fluorescent AFLP (*f*-AFLP) was used in conjunction with a capillary automated DNA sequencer, which offers a rapid route to achieving robust strain typing of brewing yeast strains. Comparative genome hybridization (CGH) to compare the

DOI: 10.1002/yea

gene content of the brewing strains to the genome-sequenced laboratory strain S288c was employed using microarray analysis. The hybridized array reveals genes common to both strains and genes that are present in the reference strain but absent in the test strain. It is known that many lager brewing strains are hybrids, making it difficult to predict their metabolic capabilities, and so the hybrid nature of the strains selected was verified by using a PCR–RFLP approach, as described by Rainieri *et al*. (2006).

In the present study, the application of chemotaxonomy at the metabolomic functional level, using metabolic footprinting to characterize yeast strains, was assessed and compared to variations observed at the genomic level, as shown by AFLP and CGH using a whole genome microarray approach. Within the National Collection of Yeast Cultures (NCYC) are a number of isolates of the *Saccharomyces sensu stricto* group which are employed in the brewing industry. Nine yeast strains were chosen for further investigation, which included two ale strains, six lager strains and the *S. cerevisiae* type strain.

Materials and methods

Yeast strains

The yeast strains used in this study are listed in Table 1. Eight of the strains are industrial isolates listed as either lager or ale strains. The ninth strain is a reference *S. cerevisiae* strain, the taxonomic type strain (NCYC 505), although this was itself originally isolated from a brewery. All strains were obtained from the National Collection of Yeast Cultures, Norwich, UK (**http://www.ncyc.co.uk**).

PCR–RFLP analysis

The hybrid nature of the yeasts in the study was examined by PCR–RFLP analysis, as detailed in Rainieri *et al*. (2006).

Growth temperature tests

The strains were grown on Yeast Extract/Malt Extract (YM; Difco, Becton, Dickenson and Co) agar, containing 0.3% w/v yeast extract, 0.3% w/v malt extract, 0.5% w/v peptone, 1% w/v glucose and 2% agar, for 3 days at both 25 °C and 37 °C.

AFLP fingerprinting

Yeast strains were grown in YM broth for 2–3 days at 25 °C on a rocker shaker. Genomic DNA was purified using the DNeasy Tissue Kit (Qiagen), following a modified version of the manufacturer's yeast protocol. Cells were collected by centrifugation at $5000 \times g$ for 10 min, resuspended in 600 µl sorbitol buffer (1 M sorbitol, 100 mM sodium EDTA, 14 mM β-mercaptoethanol) and 20 µl lyticase solution (200 U; Sigma), and incubated at 30 °C for 1.5 h to digest the cell walls. Digested cell suspensions were centrifuged at $1000 \times g$ for 10 min and the pelleted spheroplasts resuspended in 180 µl Buffer ATL (Qiagen) and 20 µl proteinase

Table 1. Strains used in this study

		Classification method		
NCYC strain	**Description**	**Historical**	**PCR–RFLP**	**Growth at 37 °C**
505	Type strain	*S. cerevisiae*	*S. cerevisiae*	*S. cerevisiae*
1187	Ale	*S. cerevisiae*	Hybrid	*S. cerevisiae*
1332	Ale	*S. cerevisiae*	Hybrid	*S. cerevisiae*
453	Lager	*S. pastorianus*	Hybrid	*S. pastorianus* or *S. bayanus*
530	Lager	*S. pastorianus*	*S. cerevisiae*	*S. cerevisiae*
680	Lager	*S. pastorianus*	Hybrid	*S. pastorianus* or *S. bayanus*
1056	Lager	*S. pastorianus*	Hybrid	*S. pastorianus* or *S. bayanus*
1324	Lager	*S. bayanus*	Hybrid	*S. pastorianus* or *S. bayanus*
2340	Lager	*S. cerevisiae*	*S. cerevisiae*	*S. cerevisiae*

Historical classification was based on traditional chemotaxonomy and represents how the strains were originally accessioned into the NCYC. The hybrid nature of strains was demonstrated by PCR–RFLP analysis, the term 'hybrid' indicating that the strain contained a mixture of *S. cerevisiae, S. bayanus* and the as-yet unidentified 'lager' genomes. Classification based on ability to grow at 37 °C differentiated between *S. cerevisiae*, which could grow at this temperature, and *S. pastorianus* and *S. bayanus*, which could not.

DOI: 10.1002/yea

K (Qiagen) prior to overnight lysis at 55 °C. The samples were treated with 4 µl RNase A (400 µg; Qiagen) for 2 min at room temperature. Following purification, DNA concentrations were determined spectrophotometrically at 260 nm and by agarose gel electrophoresis.

AFLP analysis (Keygene, The Netherlands) was carried out using the AFLP Core Reagent Kit (Invitrogen). Approximately 300 ng of each genomic DNA were digested with *Eco*RI and *Mse*I, following the manufacturer's protocol. After 2.5 h incubation at 37 °C and 10 min heat inactivation at 70 °C, *Eco*RI and *Mse*I adaptors were ligated to the restriction digest at room temperature for 2 h in a total volume of 50 µl, again following the manufacturer's protocol. Aliquots (2 µl) of a 10-fold dilution of the ligation reaction in TE buffer (10 mM Tris–HCl, 1 mM sodium EDTA, pH 7.5) were used as template in a 20 µl final volume 'non-selective' PCR with 20 pmol each of *Eco*RI-0 (5′-GACTGCGTACCAATTC-3′) and *Mse*I-0 (5′-GACGATGAGTCCTGAGTAA-3′) primers and 0.5 U Thermoprime-Plus DNA polymerase (Abgene). PCR conditions were 20 cycles of 94 °C for 30 s, 56 °C for 1 min and 72 °C for 1 min. Aliquots (2 µl) of a 50-fold dilution of this PCR in TE buffer were then re-amplified in a 20 µl final volume 'selective' PCR, using 1 pmol 6-carboxyhexachlorofluorescein (HEX)-labelled *Eco*RI-AC primer (5′-GACTGCGTACCAATTCAC-3′) and 2.3 pmol unlabelled *Mse*I-C primer (5′-GACGATGAGTCCTGAGTAAC-3′) with 1.5 mM $MgCl_2$ and 1 U Thermoprime-Plus DNA polymerase (Abgene). PCR conditions comprised an initial touchdown phase of 13 cycles of 94 °C for 30 s, 65 °C for 30 s and 72 °C for 1 min, lowering the annealing temperature by 0.7 °C at each successive cycle, followed by 23 cycles of 94 °C for 30 s, 56 °C for 30 s and 72 °C for 1 min and a final step of 72 °C for 5 min. 0.5 µl of each selective PCR was mixed with 0.125 µl GeneScan-500 ROX Size Standard (Applied Biosystems, ABI) and denatured with 9.875 µl Hi-Dye formamide (ABI) in a total volume of 11 µl at 94 °C for 3 min. For each yeast strain, 'non-selective' PCRs were performed in triplicate from the same ligation reaction and individual 'selective' PCRs carried out with these to determine the reproducibility of AFLP analysis. AFLP samples were electrophoresed in an ABI 3700 series automated DNA sequencer and the data analysed with GeneScan software (ABI). Raw GeneScan AFLP data files were converted to text files containing information of DNA fragment size and peak intensity, using an in-house programme kindly provided by Jenn Conn (John Innes Centre, Norwich, UK). The fragment size data were then converted to integer bp values and the peak intensities for each fragment size summed using an in-house programme kindly provided by Mark Reuter (IFR, Norwich, UK). AFLP profiles were trimmed so that only the size range 60–490 bp was analysed. Peak intensities were then normalized within replicates so that the total trace intensity was given a value of 1 and an average AFLP profile was calculated from replicates. AFLP fragments were then scored as 0 or 1 for their absence or presence in each profile under comparison, and a similarity index calculated using the Dice coefficient between pairs of averaged AFLP profiles (Hand *et al*., 2001). The Dice coefficient is derived from the formula: $2a/(2a + b + c)$, where a is the number of common fragments and b and c are the number of unique fragments in a given pair of AFLP profiles. Cluster analysis of the AFLP profiles was carried out based on the (1 — Dice coefficient) and dendrograms created by the unweighted pair group method using arithmetic means (UPGMA).

Comparative genome hybridization (CGH)

Microarray probes representing 6250 open reading frames (ORFs) in the *S. cerevisiae* genome were purchased from MWG and consisted of 40-mer oligonucleotides. These were printed onto Aldehyde$^+$ slides from Genetix, using an in-house arrayer built to Pat Brown's specifications (**http://cmgm.stanford.edu/pbrown/mguide/index.html**). Target DNA was extracted from strains using the protocol described by Borts *et al*. (1986), subjected to sonication (XL-2020 Sonicator®) and cup horn from Misonix) at 25% maximum power to generate random fragments of 1 or 2 kb and purified using a QIAquick® PCR purification kit (Qiagen). Sample and reference DNA was then random prime-labelled, using Cy3 and Cy5 conjugated dUTP from Amersham, and hybridized to post-processed arrays for at least 18 h at 60 °C (detailed protocols can be found at: **http://cmgm.stanford.edu/pbrown/protocols/index.html**). Following washes of 2×

DOI: 10.1002/yea

SSC, 0.14% SDS, 2× SSC, 1× SSC and 0.6× SSC, the arrays were immediately scanned and analysed using Genepix 6 and a 4000B reader from Axon Instruments. Triplicate arrays were used per yeast strain analysed. Array data were analysed using the GENCOM excel add-in, as described by Pin *et al.* (2006). Genes from each test strain that were identified as variable (i.e. absent, partially divergent or deleted) in two or more of the triplicate slide set were designated as being variable in the analysed strain compared to the S288c reference strain. Genes were then scored as 0 for variable and 1 for non-variable and a similarity index calculated using the Dice coefficient.

Metabolic footprinting: DIMS and GC–TOF–MS

Yeasts were maintained on YM agar, whose composition is described above, at 30 °C. Inocula for DIMS and GC–MS experiments were prepared from single colonies grown in 10 ml YM broth with shaking at 150 r.p.m. and 30 °C for 16 h, washed with sterile Ultrapure water, counted in a haemocytometer and inoculated into 50 ml synthetic defined minimal media supplemented with a metabolite cocktail of amino acids, organic acids and pyrimidine bases (Allen *et al.*, 2003). Cultures were inoculated at a final concentration of 5×10^5 cells/ml and incubated until stationary phase, with shaking at 200 r.p.m. and 30 °C for 20 h prior to harvesting. 1 ml samples of spent culture medium were centrifuged at $8000 \times g$ for 5 min and each supernatant divided into 200 µl aliquots for storage at −80 °C. Samples for DIMS analysis were diluted ten-fold with 30% v/v methanol in Ultrapure water containing 27 mM formic acid and centrifuged at $8000 \times g$ for 5 min immediately prior to analysis (Allen *et al.*, 2003). 450 µl samples for GC–MS analysis were spiked with internal standard (100 µl 0.18 mg/ml succinic d_4 acid) and lyophilized using a HETO VR MAXI vacuum centrifuge attached to a HETO CT/DW 60E cooling trap (Thermo Life Sciences, UK). To induce volatility and thermal stability, chemical derivatization was performed in two stages; 20 mg/ml *O*-methylhydroxyamine in pyridine (50 µl) was added and the mixture heated at 40 °C for 90 min before *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (50 µl) was added and the mixture further heated for 90 min at 40 °C. All chemicals for GC–TOF–MS were of analytical grade minimum (Sigma-Aldrich, UK). Samples for DIMS analysis were analysed by direct infusion for 2 min into a Waters electrospray Q-TOF instrument at a flow rate of 100 µl/min. Three biological and three analytical replicates were analysed. Mass spectra were exported as text files and their format changed using a Perl macro, before being combined using a Matlab (**www.mathworks.com**) macro into a *m*/*z* nominal mass vs. intensity format (both macros written in-house). The data were normalized to the highest intensity peak for each sample, and all samples combined into a single data matrix, tabulating the intensity for each mass between *m*/*z* 65 and 999. All samples were also analysed by GC–TOF–MS (Agilent 6890 gas chromatograph; LECO, Stockport, UK) coupled to LECO Pegasus III time-of-flight mass spectrometer, (LECO, Stockport, UK) as previously described (O'Hagan *et al.*, 2005). From the three biological replicates available for each strain, one biological replicate was analysed three times and the other two biological replicates once, giving a total of five analyses per strain. Raw data processing was performed using ChromaTof version 2.12 and was exported in ASCII format as a list of metabolite peaks and associated peak areas, if the peak was detected in the sample; 383 unique metabolite peaks were detected in all samples, from which 146 were identified by comparison of their mass spectra with libraries of reference compounds [authors' library and the Golm Metabolome Database (**http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/msri/gmd_msri.html**)] analysed under the same instrumental conditions (comparison based on mass spectrum). From these, 112 were definitely identified using the authors' mass spectral/retention index library (comparison based on mass spectrum and retention time), and this corresponded to 77 actual metabolites. All data were normalized to the internal standard (peak area — metabolite/peak area — internal standard) prior to subsequent multivariate analysis. This was carried out in Matlab (**http://www.mathworks.com**) for both types of data. GC–TOF–MS data were normalized to zero mean and unit variance and DIMS data to zero mean, with subsequent principal components analysis (PCA) being performed to model variation in the metabolic footprints of each strain compared to the variation associated

DOI: 10.1002/yea

within the group of strains. For DIMS data, subsequent canonical variates analysis (CVA) was performed, with each group containing all analytical replicates of a given biological replicate (hence, 9×3 groups were defined in total). For this CVA, 26 PC scores were passed on as input variables as the minimum possible for the calculations to be statistically robust. Although this approach was only 'semi-supervised' and less prone to overfitting than a fully-supervised method (in which case one group per strain would have been defined), care was taken to validate the analysis using cross-validation (validating both the PCA and CVA steps and leaving three analytical replicates out at a time). Note that this was not suitable for the GC–MS data, due to the lesser availability of analytical replicates.

Results and discussion

In this study genetic variation of nine brewing-related *Saccharomyces* strains was characterized by AFLP, PCR–RFLP and CGH. The effect of this variation on the final downstream product of gene expression, the metabolome, was examined. The discriminatory power of the various techniques was compared and conclusions drawn based on the different patterns of genetic and metabolic variation observed.

Hybrid nature and nomenclature of the brewing yeast strains used in this study

The brewing yeasts used in this study were deposited into the National Collection of Yeast Cultures between 1955 and 1989. All had been originally classified by traditional chemotaxonomic methods and their nomenclature had often changed as new methodologies were introduced.

Using the most recently described PCR–RFLP method (Rainieri *et al*., 2006), the genetic background of the brewing strains analysed was determined (Table 1). The PCR primers used were based on the *S. cerevisiae* and *S. uvarum* genomes and allowed for the amplification of three different homologous genes; *FUN14, RIP1* and *HIS3* from *S. cerevisiae, S. uvarum* and *S. bayanus* genomes. PCR–RFLP analysis was used to differentiate between *S. uvarum* and *S. bayanus* genomes.

Interestingly, the two ale strains designated as *S. cerevisiae* were found to contain both *S. cerevisiae* and *S. bayanus* RFLPs. This suggested that they should be designated as *S. pastorianus* under the Rainieri *et al*. (2006) proposal that the *S. pastorianus* name should be used for the multiple genetic lines that contain the *S. cerevisiae* genome, viz. the *S. cerevisiae*/*S. bayanus* hybrid line and the *S. cerevisiae*/*S. bayanus*/*S. uvarum* hybrid line.

Four of the six lager strains (NCYC 453, NCYC 680, NCYC 1056 and NCYC 1324) were correctly (according to Rainieri *et al*., 2006) designated *S. pastorianus*, as they were also found to contain both *S. cerevisiae* and *S. bayanus* RFLPs. The remaining two lager strains (NCYC 530 and NCYC 2340) were found to contain only a *S. cerevisiae* background, despite one (NCYC 530) having been deposited originally as *S. pastorianus*.

Ability of ale strains to grow at 37 °C

The capacity for growth at 37 °C is currently regarded as an additional differentiating factor, allowing for taxonomic separation of *S. cerevisiae*, which can grow at 37 °C, from strains of *S. pastorianus* or *S. bayanus*, which do not have this ability (Vaughan-Martini and Martini, 1993; Walsh and Martin, 1977). The two ale strains analysed grew at 37 °C (Table 1), suggesting that they should be classified as *S. cerevisiae*, which is contrary to the hybrid nature determined by PCR–RFLP but consistent with the historical classification. The two lager strains shown by PCR–RFLP to be pure *S. cerevisiae* strains (NCYC 530 and NCYC 2340) were also able to grow at 37 °C, which is in agreement with previous temperature differentiation tests. Likewise, all lager strains shown to be hybrids by PCR–RFLP were unable to grow at 37 °C. These results highlight the limitations of such techniques for discrimination amongst industrial strains and more precise methodologies are discussed below.

Genetic fingerprinting of ale and lager brewing yeasts

AFLP analysis was used to give a 'snapshot' of DNA sequence variation across the whole genome. The banding patterns are shown in Figure 1 and the resulting cluster analysis based on the Dice coefficient in Figure 2. Greater genetic diversity is shown by larger 1-Dice coefficient values. AFLP analysis separated the yeast strains into two main clusters. The lager strains NCYC 680, NCYC 1056,

DOI: 10.1002/yea

NCYC 1324 and NCYC 453 grouped together, with NCYC 453 and 1324 being most genetically similar. The lager strain NCYC 530 and the *S. cerevisiae* type strain NCYC 505 clustered with the ale strain NCYC 1332. The ale strain NCYC 1187 and the lager strain NCYC 2340 were shown to be the most genetically diverse when compared to all the other strains studied. Although AFLP gave good strain discrimination, the observed groupings did not correlate precisely with industrial classifications.

CGH of ale and lager brewing yeasts

A microarray comprising of 6250 ORFs in the *S. cerevisiae* S288c genome was used to compare the whole genome gene content diversity among the nine yeast strains. The GENCOM algorithm (Pearson *et al*., 2003; Pin *et al*., 2006) was used to analyse the microarray data by comparing the ratios of the fluorescence signals obtained to predict whether

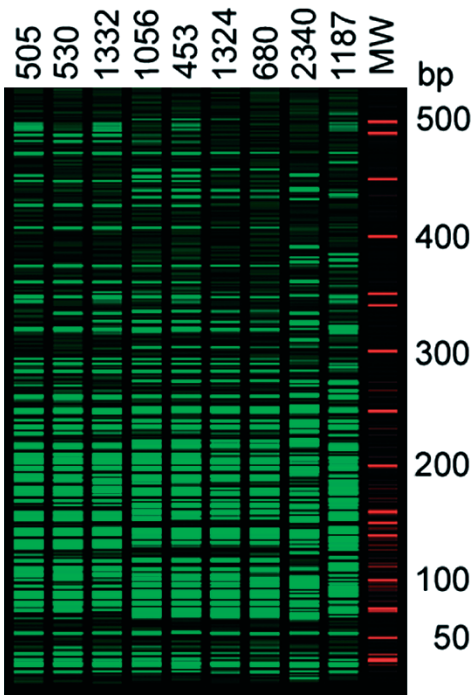

Figure 1. Computer-generated gel image of AFLP traces from the nine brewing yeast strains. Fluorescently HEX-labelled AFLP traces are shown in green, along with one of the internal ROX-500 molecular weight marker traces shown in red. The image was created using Genographer version 1.6.0 freeware (Benham JJ; **http://hordeum.oscs.montana.edu/genographer/**)

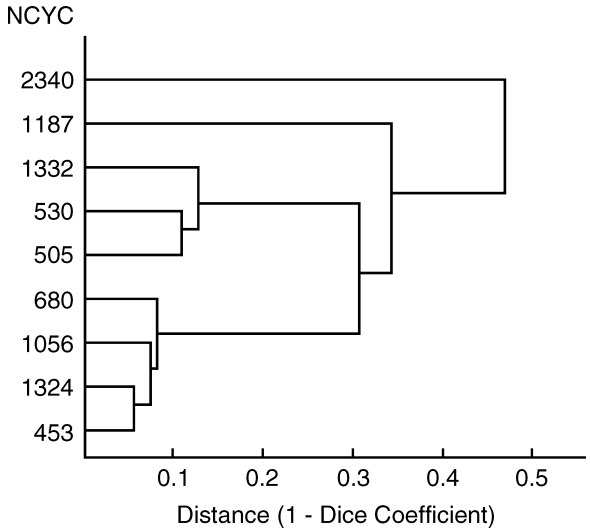

Figure 2. Phylogenetic tree derived from cluster analysis of AFLP patterns from the nine brewing yeast strains

or not a gene is variable. A variable gene could be absent from the genome, partially deleted or divergent, to the extent that it gives a significantly lower hybridization signal than the control strain. The hybridized array revealed genes common to both strains and genes that are present in the reference strain but absent in the test strain. This method cannot, of course, detect genes present in the test strain but missing in the reference strain. Strains found to be similar in terms of having similar genes absent compared to the reference strain (see Supplementary Table 1), fell into two main groups. Group 1 included the ale strain NCYC 1187 and the lager strains NCYC 530, NCYC 680, NCYC 1056 and NCYC 2340. Group 2 included the lager strains NCYC 453 and NCYC 1324, the ale strain NCYC 1332 and the reference strain NCYC 505. These results did not correlate with: (a) historical strain information; (b) genetic variation as identified by AFLP analysis; or (c) metabolic profiling (see below). This may be because the reference strain S288c is a laboratory strain and may be missing genes associated with brewing (e.g. for flocculation, head formation, maltose metabolism) that would have aided in the differentiation between brewing strain types. Many of the strains analysed were identified as hybrids and thus, without comparing them to reference *S. pastorianus* and *S. bayanus* whole genome arrays, it is hard to interpret the data obtained, due to the complexity of

DOI: 10.1002/yea

the biological pathways involved. Nevertheless, we conclude that currently available methods of CGH analysis failed to provide meaningful discrimination in an industrial strain context.

We did, however, identify a number of variable genes that are associated with carbohydrate metabolism in brewing, such as *ADH7, AAD3, AAD15* and *AAD16* (alcohol metabolism), *ALD2* (aldehyde metabolism), *HXT15* and *HXT16* (glucose metabolism), *MAL33* (maltose metabolism) and *GDB1* and *GPH1* (glycogen metabolism). A number of asparaginases were also identified to be variable (*ASP3-1, ASP3-2, ASP3-3* and *ASP3-4*) and the presence or absence of these genes may be of future discriminatory importance, as it is known that asparagine is a key amino acid present in wort (Jones and Pierce, 1964).

Metabolic footprinting of ale and lager brewing yeasts

GC–TOF–MS and DIMS detected a wide variety of metabolites in the metabolic footprint, including amino and organic acids, carbohydrates, lipids, alcohols and phosphorylated compounds (see Supplementary Tables 2 and 3), demonstrating the richness of biological information that can be obtained using these approaches.

The DIMS data were analysed by PCA (Figure 3a), the scores of which were then further analysed by CVA in a fully validated model (Figure 3b). The scores plots showed similar clustering but the combination of techniques allowed further interpretation to be carried out. Both plots show that there was a clear grouping of the ale strains, NCYC 1187 and NCYC 1332. The lager strains NCYC 680 and NCYC 1056 formed a cluster with two other lager strains, NCYC 453 and NCYC 1324 (although these two sets of strains are separated on PC2, which represents 10.1% of the variance, they are very close on PC1, which represents 73% of the variance; they are also close on the CVA plot). The reference *S. cerevisiae* type strain NCYC 505 exhibited more variability between biological replicates than other strains, and was positioned between the ale strains and the group of lager strains described above. The lager strain NCYC 2340 was overall similar to the group of lager strains but was found to differ at a number of specific *m/z* values (i.e. higher intensities at *m/z* 86 as well as 99, 113, 72, and lower intensities at *m/z* 84 and 130), which caused it to be separated in the CVA on CV2, and in the PCA on the fifth PC axis (1.9% of the variance). The lager strain NCYC 530 distinguished itself from all others on CV2 (and it did this only slightly on the second PC axis), and this was because of extremely low responses at *m/z* 110 and 156 and a quite low response at *m/z* 80. The results obtained with DIMS were extremely encouraging in their ability to explore the similarities between strains, but an inherent limitation of the technique is in terms of interpretability. Chromatography–mass spectrometry technologies are thus required for metabolic identification.

PCA of GC–TOF–MS data (Figure 4) also clearly separated the ale strains, NCYC 1187 and NCYC 1332, from the others. Metabolites showing

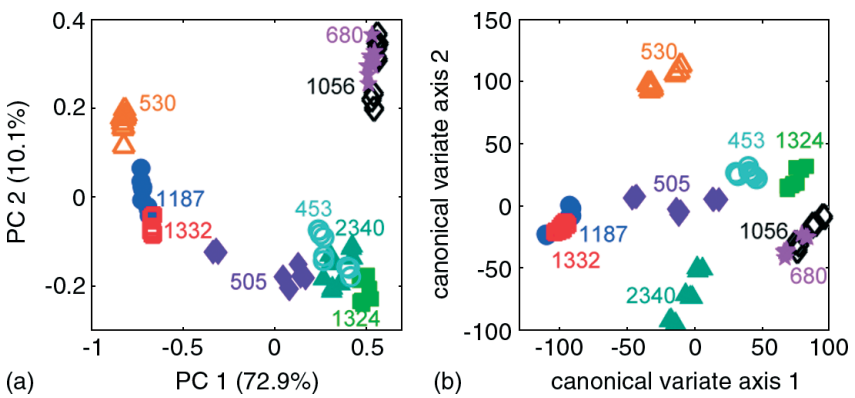

Figure 3. Metabolic footprinting of brewing yeast strains by DIMS, (a) analysed by PCA and (b) validated by subsequent semi-supervised CVA. NCYC 505 is the type strain of *S. cerevisiae*, NCYC 1187 and NCYC 1332 are *S. cerevisiae* ale strains, while all others are lager strains of *S. cerevisiae, S. pastorianus* or *S. bayanus*

DOI: 10.1002/yea

concentration-related differences between strains NCYC 1187 and NCYC 1332, and the other strains included trehalose and organic/fatty acids, such as octadecanoic, lactic and *trans*-aconitic. Indeed, trehalose was below the level of detection in the type strain and all the lager strains tested and only found in the two ale strains (Figure 5 and see Supplementary Table 4). Trehalose acts as a protectant that contributes to the survival of yeast under various stressful conditions (Fernandes, 2005; Vanlaere, 1989). It has been found to accumulate in high amounts in ethanol-tolerant mutant sake yeast (Ogawa *et al*., 2000) and as such is a beneficial characteristic for brewing yeast. The results presented here suggest that this metabolite may be useful in discrimination between brewing strains. The type strain NCYC 505 and the lager strain NCYC 530 can be seen to separate on PC2 away from the other strains with GC–TOF–MS data. The lager strains NCYC 1056 and NCYC 680 clustered with the lager strains NCYC 453 and NCYC 1324. A number of amino acids (valine, lysine, alanine and methionine), maltose and pyruvic acid showed concentration-related differences for these four strains relative to other strains, based on the GC–TOF–MS data (see Supplementary Table 3). The lager strain NCYC 2340 was isolated again; however, there was only a small separation in PC4 between NCYC 2340 and two other lager strains, NCYC 453 and NCYC 1324. Of interest is the discriminatory power of these techniques; GC–TOF–MS was simply analysed by unsupervised PCA and allows for definitive metabolite identification with a throughput of 80 samples/day, whereas DIMS allowed for rapid screening (2 min) with good separation by PCA or CVA and a higher-throughput approach compared to GC–TOF–MS (15 min), but lacked the ability for specific metabolites to be identified.

The metabolic footprinting clustering patterns obtained from multivariate data analysis of both DIMS and GC–TOF–MS data were similar. Significantly, in both analyses the ale strains (NCYC 1187 and NCYC 1332) formed a defined cluster, separated from the lager strains. Generally the lager strains NCYC 453, NCYC 1324, NCYC 680 and NCYC 1056 group together. However, the lager strain NCYC 530 has a different metabolic profile in both DIMS and GC–TOF–MS compared to all of the other analysed strains, but is most metabolically similar to the type strain NCYC 505. The lager strain NCYC 2340 shows the most variance between the two metabolic footprinting methods, having a metabolic profile most similar to the main cluster of lager strains (NCYC 453,

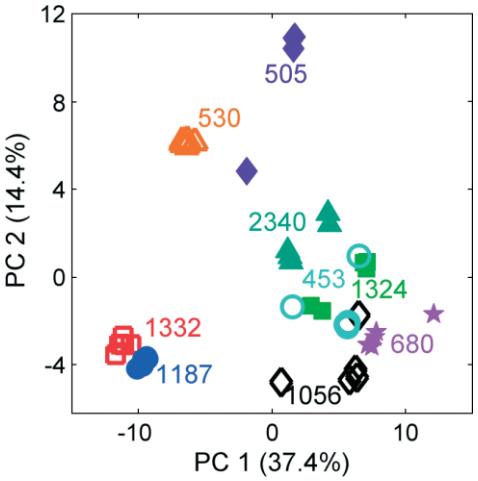

Figure 4. Metabolic footprinting of brewing yeast strains by GC–TOF–MS, analysed by PCA. NCYC 505 is the type strain of *S. cerevisiae*, NCYC 1187 and NCYC 1332 are *S. cerevisiae* ale strains, while all others are lager strains of *S. cerevisiae, S. pastorianus* or *S. bayanus*

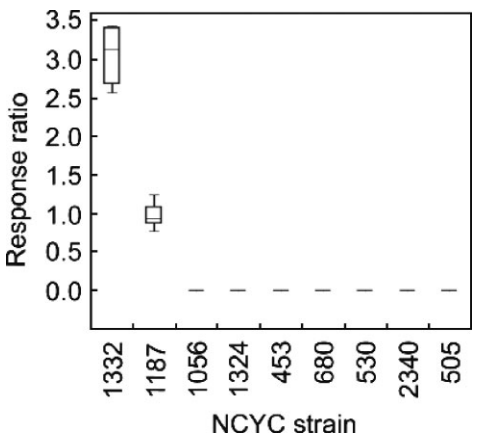

Figure 5. 'Box-and-whisker' plot of relative trehalose levels in GC–TOF–MS samples from the nine brewing yeast strains. The response ratio is the measure of trehalose signal relative to the internal standard in each sample. The lower and upper edges of the 'box' are the 25th and 75th percentiles of the samples, with the distance between them being the inter-quartile range. The sample median is represented by the line inside the 'box' and its position away from the centre represents the degree of skewness in the data. The 'whiskers' extending from the upper and lower edges indicate the maximum and minimum sample values, respectively

NCYC 680, NCYC 1056 and NCYC 1324) in the GC–TOF–MS PCA, whereas classification based on DIMS data suggests a more diverse profile compared to the other strains analysed.

When performing metabolic footprinting, care must be taken to ensure that any differences observed between strains are not growth stage-related. This is normally achieved by inoculating cultures with the same number of cells and incubating for a sufficient period of time following entry into stationary phase before harvesting the culture medium (Allen *et al*., 2003). In the present study, the same numbers of cells were used as starting inocula but variability in cell size may have resulted in variation in starting OD_{600nm} values and subsequent final OD_{600nm} values (Table 2). Differences in growth rate between strains also could not be ruled out, but when the final OD values were compared with the PCA and CVA plots for both the DIMS and GC–TOF–MS data, there was no trend relating PC1, PC2, CV1 and CV2 scores to growth (Figure 6), thus the discrimination between strains was not simply growth-related.

Similarly, the estimated number of elapsed generations did not correlate with PC1, PC2, CV1 and CV2 scores (see Supplementary Figure 1) On the contrary, many of the strains clustering metabolically showed large differences in their growth pattern, e.g. the NCYC 1187/NCYC 1332 and NCYC 680/NCYC 1056 metabolic clusters. In addition,

Table 2. Growth of strains for metabolic footprinting

NCYC strain	Starting $OD_{600\ nm}$	Final $OD_{600\ nm}$	Estimated number of generations
530	0.18	7.9	5.4
2340	0.12	7.0	5.8
1332	0.08	6.2	6.2
505	0.08	5.0	6.0
1056	0.09	4.9	5.7
1187	0.12	4.9	5.3
453	0.14	4.8	5.1
1324	0.08	4.8	5.9
680	0.09	3.5	5.2

Washed inocula were prepared as described in Materials and methods and added at an initial concentration of 5×10^5 cells/ml to the defined metabolite cocktail-containing medium of Allen *et al.* (2003). Cultures were incubated at 30 °C, 200 r.p.m. for 20 h prior to harvesting.

many of the discriminatory metabolites between strains in the PCA were not present in the 0 h post-inoculated growth medium and hence were secreted from the cells. These differences were therefore not due to altered consumption rates of growth medium components. Undoubtedly, growth rate and growth phase will have an influence on strain separation in PCA space, but this is considered to be negligible in the present study because of the trends in the data described above.

The method is being developed for rapid, high-throughput analysis of samples and hence requires to be robust with respect to differences in growth

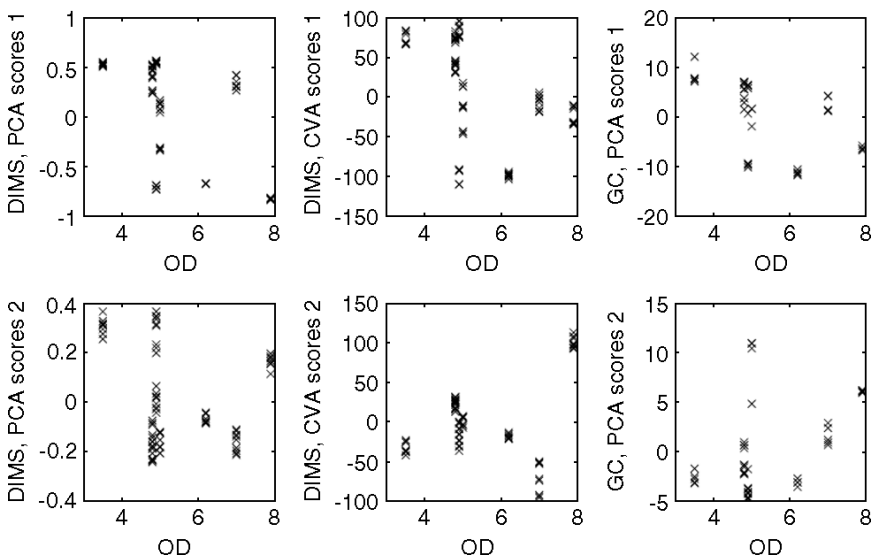

Figure 6. Lack of relationship between growth, as expressed as final culture OD_{600nm}, and PCA/CVA scores from metabolic footprinting

DOI: 10.1002/yea

rate between strains within certain limits. With this in mind, different growth media, some more minimal than that used in the present study, are currently being examined for their usefulness in metabolic footprinting, but the metabolite cocktail-containing medium used by Allen *et al*. (2003) has already been optimized for probing 'overflow' metabolism in yeast.

Metabolic footprinting clearly separates ale and lager strains

The analysis of metabolic footprints by GC–TOF–MS and DIMS, and subsequent multivariate analysis, showed a similar but not identical pattern of diversity, as evident from AFLP analysis. Of particular interest was the clear separation of ale and lager strains by metabolic footprinting, as would be expected biologically.

Both metabolic footprinting and AFLP analysis separated the lager strains NCYC 453, NCYC 680, NCYC 1056 and NCYC 1324 from all of the other strains in the study, showing that they are the least diverse of the strains analysed. Despite originating from geographically diverse regions (UK, Japan and Czech Republic), these strains are likely to have been derived originally from a similar brewing strain. These strains were also all shown to be hybrid in nature by PCR–RFLP analysis (containing a mixture of *S. cerevisiae, S. bayanus* and lager genomes) and were unable to grow at elevated temperature (37 °C).

The lager strain NCYC 2340 (originating from Belgium) was clearly both genetically and metabolically divergent to the other strains analysed, as it showed a distinctive metabolic profile (using DIMS data) and genetic profile (AFLP data). NCYC 530 is clearly more metabolically diverse than any other strain analysed, yet using AFLP analysis it was shown to be genetically similar to NCYC 505 and NCYC 1332. This result shows the discriminatory power of metabolic footprinting compared to genomic approaches and its applicability to a highly discriminatory chemotaxonomic methodology. This is also shown for the ale strains NCYC 1187 and NCYC 1332 below.

Importantly, on the basis of their genetic data, the ale strains NCYC 1187 and NCYC 1332 clustered with different groups of lager strains, rather than with each other, but metabolically they cluster together in multivariate space. Both strains were from UK sources and both were shown to be hybrids. Their AFLP patterns indicated that they are genetically diverse in the context of this strain set but they clearly produce similar metabolic profiles.

Concluding remarks

It is known that changes in the metabolome are amplified relative to the proteome and transcriptome (Kell and Westerhoff, 1986) and therefore provide a more sensitive system to determine the effects of genomic differences in many different biological systems (Allen *et al*., 2003; Fiehn *et al*., 2000; Kaderbhai *et al*., 2003). Analysis of the endometabolomes of different strains of *Escherichia coli* by two-dimensional high-performance thin-layer chromatography showed a wide range of biodiversity that correlated well with the genetic relationships among strains (Maharjan and Ferenci, 2005). Similary, metabolomic analysis of whole cells from different *Candida* species by nuclear magnetic resonance spectroscopy provided a rapid method for identification and intra-specific strain discrimination (Himmelreich *et al*., 2003, 2005). The results presented here show that the exometabolome may be used to discriminate and classify industrial yeast strains and provide a new approach for chemotaxonomy. Several discriminatory metabolites were identified. In particular, metabolic footprinting had the ability to group the ale strains, which could not be achieved by AFLP and gene content analysis. The two metabolomic methods employed gave similar results. While the groupings obtained do not correlate well with groupings observed by genetic methods, they are more discriminatory and more readily interpreted. However, because of the observed discrepancies in the metabolic and genetic data, a combination of both approaches should be used for a fully comprehensive strain characterization. The ale strains show divergence at the genotypic level despite conservation at the phenotypic level, suggesting that genetic divergence does not always lead to metabolic divergence. We conclude that there is an exciting possibility of employing the exometabolome and metabolic footprinting for high-throughput yeast chemotaxonomy and to interpret the results in terms of commercially

DOI: 10.1002/yea

important traits, as well as in terms of evolution of novel pathways and novel biological variation.

Acknowledgements

We would like to thank Sandra Rainieri for her help in classifying the yeast strains used in this study. The work was supported by the Biotechnology and Biological Sciences Research Council.

References

Allen J, Davey HM, Broadhurst D, *et al*. 2003. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnol* **21**: 692–696.

Allen J, Davey HM, Broadhurst D, *et al*. 2004. Discrimination of modes of action of antifungal substances by use of metabolic footprinting. *Appl Environ Microbiol* **70**: 6157–6165.

Azumi M, Goto-Yamamoto N. 2001. AFLP analysis of type strains and laboratory and industrial strains of *Saccharomyces sensu stricto* and its application to phenetic clustering. *Yeast* **18**: 1145–1154.

Borts RH, Lichten M, Haber JE. 1986. Analysis of meiosis-defective mutations in yeast by physical monitoring of recombination. *Genetics* **113**: 551–567.

de Barros Lopes M, Bellon JR, Shirley NJ, Ganter PF. 2002. Evidence for multiple interspecific hybridization in *Saccharomyces sensu stricto* species. *FEMS Yeast Res* **1**: 323–331.

Deak T, Beuchat LR (eds). 1996. *Handbook of Food Spoilage Yeasts*. CRC Press: Boca Raton, FL, USA.

Dunn WB, Bailey NJC, Johnson HE. 2005. Measuring the metabolome: current analytical technologies. *Analyst* **130**: 606–625.

Fernandes PMB. 2005. How does yeast respond to pressure? *Braz J Med Biol Res* **38**: 1239–1245.

Fiehn O, Kopka J, Dormann P, *et al*. 2000. Metabolite profiling for plant functional genomics. *Nature Biotechnol* **18**: 1157–1161.

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. *TIBTECH* **22**: 245–252.

Hand DJ, Smyth P, Mannila H (eds). 2001. *Principles of Data Mining*, 1st edn. MIT Press: Cambridge, MA, USA.

Hansen J, Kiellandbrandt MC. 1994. *Saccharomyces carlsbergensis* contains two functional *MET2* alleles similar to homologs from *Saccharomyces cerevisiae* and *S. monacensis*. *Gene* **140**: 33–40.

Himmelreich U, Somorjai RL, Dolenko B, Daniel HM, Sorrell TC. 2005. A rapid screening test to distinguish between *Candida albicans* and *Candida dubliniensis* using NMR spectroscopy. *FEMS Microbiol Lett* **251**: 327–332.

Himmelreich U, Somorjai RL, Dolenko B, *et al*. 2003. Rapid identification of *Candida* species by using nuclear magnetic resonance spectroscopy and a statistical classification strategy. *Appl Environ Microbiol* **69**: 4566–4574.

Howell KS, Cozzolino D, Bartowsky EJ, Fleet GH, Henschke PA. 2006. Metabolic profiling as a tool for revealing *Saccharomyces* interactions during wine fermentation. *FEMS Yeast Res* **6**: 91–101.

Jones M, Pierce JS. 1964. Absorption of amino acids from wort by yeasts. *J Inst Brew* **70**: 307–315.

Kaderbhai NN, Broadhurst DI, Ellis DI, Goodacre R, Kell DB. 2003. Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. *Comp Funct Genom* **4**: 376–391.

Kell DB. 2004. Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* **7**: 296–307.

Kell DB. 2006. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov Today* **11**: 1085–1092.

Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG. 2005. Metabolic footprinting and systems biology: the medium is the message. *Nature Rev Microbiol* **3**: 557–565.

Kell DB, Westerhoff HV. 1986. Metabolic control-theory — its role in microbiology and biotechnology. *FEMS Microbiol Rev* **39**: 305–320.

Koeleman JGM, Stoof J, Biesmans DJ, *et al*. 1998. Comparison of amplified ribosomal DNA restriction analysis, random amplified polymorphic DNA analysis, and amplified fragment length polymorphism fingerprinting for identification of *Acinetobacter* genomic species and typing of *Acinetobacter baumannii*. *J Clin Microbiol* **36**: 2522–2529.

Maharjan RP, Ferenci T. 2003. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Anal Biochem* **313**: 145–154.

Maharjan RP, Ferenci T. 2005. Metabolomic diversity in the species *Escherichia coli* and its relationship to genetic population structure. *Metabolomics* **1**: 235–242.

Nguyen HV, Lepingle A, Gaillardin C. 2000. Molecular typing demonstrates homogeneity of *Saccharomyces uvarum* strains and reveals the existence of hybrids between *S. uvarum* and *S. cerevisiae* including the *S. bayanus* type strain CBS 380. *System Appl Microbiol* **23**: 71–85.

Ogawa Y, Nitta A, Uchiyama H, *et al*. 2000. Tolerance mechanism of the ethanol-tolerant mutant of sake yeast. *J Biosci Bioeng* **90**: 313–320.

Oliver SG, Winson MK, Kell DB, Baganz F. 1998. Systematic functional analysis of the yeast genome. *TIBTECH* **16**: 373–378 [erratum **16**: 447].

Pearson BM, Pin C, Wright J, I'Anson K, Humphrey T, Wells JM. 2003. Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Lett* **554**: 224–230.

Pin C, Reuter M, Pearson B, *et al*. 2006. Comparison of different approaches for comparative genetic analysis using microarray hybridization. *Appl Microbiol Biotechnol* **72**: 852–859.

Raamsdonk LM, Teusink B, Broadhurst D, *et al*. 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnol* **19**: 45–50.

Rainieri S, Kodama Y, Kaneko Y, *et al*. 2006. Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome. *Appl Environ Microbiol* **72**: 3968–3974.

Tornai-Lehoczki J, Dlauchy D. 2000. Delimination of brewing yeast strains using different molecular techniques. *Int J Food Microbiol* **62**: 37–45.

Urbanczyk-Wochniak E, Luedemann A, Kopka J, *et al*. 2003. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep* **4**: 989–993.

Vanlaere A. 1989. Trehalose, reserve and or stress metabolite. *FEMS Microbiol Rev* **63**: 201–210.

Vaughan-Martini A, Kurtzman CP. 1985. Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces sensu stricto*. *Int J System Bacteriol* **35**: 508–511.

Vaughan-Martini A, Martini A. 1993. A taxonomic key for the genus *Saccharomyces*. *System Appl Microbiol* **16**: 113–119.

Villas-Bôas SG, Højer-Pedersen J, Åkesson M, Smedsgaard J, Nielsen J. 2005. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **22**: 1155–1169.

Villas-Bôas SG, Noel S, Lane GA, Attwood G, Cookson A. 2006. Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media. *Anal Biochem* **349**: 297–305.

Walsh RM, Martin PA. 1977. Growth of *Saccharomyces cerevisiae* and *Saccharomyces uvarum* in a temperature-gradient incubator. *J Inst Brew* **83**: 169–172.

Wightman P, Quain DE, Meaden PG. 1996. Analysis of production brewing strains of yeast by DNA fingerprinting. *Lett Appl Microbiol* **22**: 90–94.

Yamagishi H, Otsuta Y, Funahashi W, Ogata T, Sakai K. 1999. Differentiation between brewing and non-brewing yeasts using a combination of PCR and RFLP. *J Appl Microbiol* **86**: 505–513.