Systematic functional analysis of the yeast genome

Stephen G. Oliver, Michael K. Winson, Douglas B. Kell and Frank Baganz

S. G. Oliver (steve.oliver@umist.ac.uk) and F. Baganz are in the Department of Biomolecular Sciences, UMIST, PO Box 88, Manchester M60 1QD, UK.
M. K. Winson and D.B. Kell (mkw,dbk@aber.ac.uk) are in the Institute of Biological Sciences, University of Wales, Aberystwyth SY23 3DD, UK.

The genome sequence of the yeast, *Saccharomyces cerevisiae*, has provided the first complete inventory of the working parts of a eukaryotic cell. The challenge is now to discover what each of the gene products does and how they interact in a living yeast cell. Systematic and comprehensive approaches to the elucidation of yeast gene function are discussed and the prospects for the functional genomics of eukaryotic organisms evaluated.

The new field of functional genomics[1] presents yeast researchers, in particular, with new responsibilities and opportunities. The responsibility is to elucidate the function of each and every one of the almost 4000 novel protein-encoding genes discovered by the *Saccharomyces cerevisiae* Genome Sequencing Project[2,3]. The opportunity is to determine how all yeast genes, both those that were discovered by classical ('function first') genetics and those that were revealed by the complete genome sequence, interact to allow this simple eukaryotic cell to grow, divide, develop and respond to environmental changes. Thus functional genomics should not only provide essential information about the role of novel genes, it should also throw new light on the contributions made by the 'old' genes. If this holistic, or fully integrative, view of the yeast cell can be obtained, it should provide an important 'navigational aid'[4] to guide our studies of more complex genomes, such as those of humans, crop plants and farm animals.

Functional genomics requires the development of analytical strategies that are comprehensive and hierarchical. Comprehensive, because we aim to uncover the action and interaction of all of the genes in a given species. Hierarchical, because this daunting task is possible only if we find ways of grouping genes of related function in order to limit the total number of experiments to be performed. Having achieved such a grouping, we can then construct smaller and smaller subgroups, proceeding down the hierarchies of analysis, achieving a closer and closer approximation to the functions of each novel gene[4-6]. At the highest level of this hierarchy, the comprehensive nature of the analytical methods employed requires that special care is taken over the design of experiments. In classical or 'function first' genetics, mutants are isolated that show a specific phenotype, e.g. a requirement for nutrient X or the failure to produce product Y. Once the gene defined by the mutation has been identified, it is known without a doubt that it controls the synthesis of X or Y within the living cell, although it may be far from obvious how the gene does this. On the other hand, in functional genomics, we look at the activity of *all* of the genes. Because not all of them will be relevant to the property under study, it is necessary to design experiments that are as precise as possible in the biological questions that they pose. This means that we need to define the physiological or developmental state of the cells much more accurately than we have generally been accustomed to doing. Moreover, we need to devise ways of sampling gene products (be they transcripts or proteins) and metabolic intermediates in ways which ensure that they accurately reflect, both qualitatively and quantitatively, their status within the living cell. This problem is particularly acute when studying the activity of pathogens in their hosts.

Generation of specific deletion mutants

In order to determine the function of novel yeast genes using a systematic and hierarchical approach, a large European research network, called EUROFAN (for European Functional Analysis Network) has been established[7], and parallel activities are under way in Canada, Japan and the USA. The first aim of these projects is the production of a complete set of 6000 single-gene deletion mutants covering all of the open reading-frames (ORFs) revealed by the complete *Saccharomyces cerevisiae* genome sequence. Such a mutant collection will represent a major resource, not only for the functional analysis of the yeast genome but should also facilitate analysis of higher genomes by permitting their 'functional mapping' onto that of yeast through transcomplementation experiments[4,6].

These deletions are being generated by a PCR-mediated gene replacement method which relies on the high efficiency and accuracy of mitotic recombination in *S. cerevisiae*. This technique was first described for PCR-generated DNA molecules consisting of the *S. cerevisiae HIS3* gene flanked by 30-50 bp of DNA homologous to the target locus[8]. However, the use of homologous nutritional markers (e.g. *HIS3* or *TRP1*) in gene replacement experiments resulted in a rather low efficiency of correct integration, particularly in strains that do not carry a complete deletion of the marker locus[9]. Large improvements were achieved when kanMX, a completely heterologous dominant resistance marker, was used as a selectable module[10]. The heterology of the kanMX module means that correct disruption of target ORFs with PCR-generated kanMX cassettes containing 35 bp flanking extensions, homologous to yeast chromosomal DNA, can be carried out with a very low level of illegitimate integration[10]. Successful targeting of these PCR products is dependent on perfect homology between the short flanking regions and the target locus. In order to permit gene replacement in any *S. cerevisiae* strain, including those used in industry, an alternative PCR method was developed[11,12] in which a gene disruption cassette with long flanking homology regions is generated in a nested PCR reaction. The current status of the deletion analysis of of novel ORFs by the EUROFAN research network[7] is shown in Table 1.

The transcriptome

A powerful tool in the elucidation of gene function is the quantitative analysis of expression levels of a novel gene under a variety of physiological or developmental conditions, compared with the expression patterns of functionally characterised genes. One approach to the efficient analysis of gene expression on a genome-wide scale is the use of SAGE (serial analysis of gene expression) technology[13], which was employed by Velculescu and co-workers[14] to determine the complete set of yeast genes expressed under a given set of conditions (the 'transcriptome'). In their initial studies, they were able to detect 4665 genes (ca. 75% of the predicted protein-encoding ORFs) with most genes being expressed at a low level. Kal[15] has used the SAGE technology to compare yeast gene transcription in cells grown on oleate and cells grown on glucose, and found that peroxisomal activity is induced by growth on oleate; the 35 genes showing the highest oleate:glucose transcript ratio included many known to encode peroxisomal proteins, but 15 novel genes were also identified as being likely to be associated with peroxisomal metabolism.

An alternative approach to SAGE is the use of hybridisation-array technologies[16,17] for the high-throughput analysis of the transcriptome. DeRisi and co-workers[18] have applied DNA microarrays containing almost every yeast ORF to carry out a comprehensive study of gene expression during the metabolic shift from fermentation (involving the production of ethanol from glucose via glycolysis) to respiration (involving the aerobic utilisation of the ethanol once glucose has been exhausted) on a genomic scale. They found marked changes in the global pattern of gene expression for ca. 35% of the ca. 6000 putative protein-encoding genes as glucose was progressively depleted from the medium. About half of these differentially expressed genes were previously uncharacterised, and so the response to the shift from fermentation to respiration allowed them to be grouped together at the start of the hierarchy of analysis. The expression profiles observed for genes with known functions provided insights into the cell's response to a changing environment. For example, several classes of genes, such as those involved in the TCA/glyoxylate cycle and carbohydrate storage, were coordinately induced by glucose exhaustion, whereas genes involved in protein synthesis exhibited a decrease in expression.

There are other ways to group genes on the basis of similarities in their expression patterns. When previously characterised members of such groups share important similarities in their functions, common regulatory mechanisms may be inferred and the likely function of novel genes in the group may be guessed at. For instance, DeRisi *et al.* [18] showed that the same DNA microarrays may be used to identify genes whose expression was affected by deletion of the gene encoding the transcriptional co-repressor Tup1p or overexpression of the transcriptional activator gene, *YAP1*. Their results demonstrate the feasibility and utility of the DNA microarray methodology in the analysis of the transcriptome and indicate that this method should enable the classification of all novel genes according to their expression patterns.

The DNA microarrays used by De Risi *et al.*[18] consisted of PCR products produced using primers complementary to the ends of ORFs. An alternative is to use oligonucleotides synthesised *in situ* on a chip using photolithography. (Both these techniques are reviewed by Hoheisel[19].) In a first application of this technique to the yeast transcriptome, Wodicka and coworkers[20] have compared the transcripts present in yeast cells grown on a rich, glucose-containing medium with those of cells grown on a minimal medium containing the same carbon source. Whilst convincing evidence is provided for the potential of the oligo-chip technology, there is little of biological interest that can be gleaned from this particular experiment due to the large number of variables involved.

The Proteome

Another route to the elucidation of gene function is the analysis of all yeast proteins synthesised under a given set of conditions, the so-called 'proteome'[21]. The yeast proteome is being defined by two-dimensional gel electrophoreses[22] and mass spectrometry, to identify the proteins within the spots on the gel[23,24]. By taking full advantage of the complete genome sequence, it should be possible to determine a single protein fragment of unique mass in order to identify a yeast protein. The use of rapid and accurate mass spectrometric methods should then allow us to monitor changes in the protein maps of yeast strains grown under different physiological conditions, as well as to identify the modification of proteins as a result of the deletion or overexpression of a novel gene. In the long term, this information will enable the identification of all components of protein complexes, isolated by either biochemical separations or immunoprecipitation, and thus create a map of all interactions between the yeast proteins, which can be correlated with *in vivo* data obtained with the yeast two-hybrid system. Such a comprehensive two-hybrid approach has already generated a protein linkage map for the *E. coli* bacteriophage T7 genome, which encodes 55 proteins[25]. It is known that the 2-hybrid approach can sometimes generate false positives, and a highly selective two-hybrid procedure adapted for exhaustive screens of the yeast genome is currently being developed by Fromont-Racine and co-workers[26].

Quantitative phenotypic analysis

One reason for taking a quantitative approach to the analysis of gene function is the growing number of genes, found by systematic genome sequencing projects, that have homologs of unknown function in a number of species, but which had not been discovered by classical molecular genetics. It is possible that these genes have been missed because quantitative, rather than qualitative, data are required to reveal their phenotypic effect. Another reason to consider a quantitative approach to phenotypic analysis is the high level of redundancy that is apparent in the yeast genome. If a particular gene is a member of a paralogous set of identical (or nearly identical) genes then, provided that they are all regulated in a similar manner and their products are targeted to the same cellular location, the contribution which an individual member of the set makes to the phenotype will, necessarily, be some fraction of the whole.

A conceptual and mathematical framework for the quantitative analysis of gene function is provided by metabolic control analysis (MCA) as pioneered by Kacser and Burns[27] and Heinrich and Rapoport [28]. The central mathematical device of MCA is the Control Coeffcient (C^J_E), a dimensionless number expressing the relative change in some variable, J (a flux or metabolite concentration) caused by a small change in the activity of some effector, E (usually, but not necessarily, an enzyme). Flux-control coefficients express the proportional contribution which any given enzyme makes to the flux through a metabolic pathway. The sum of all the flux control coefficients ($\Sigma\, C^J_E$) relevant to a given pathway is 1. Thus, in linear pathways, individual flux-control coefficients will normally have values between 0 (no control) and 1 (completely rate-determining enzyme) but, in branched pathways, values outside this range (i.e. < 0 or > 1) can occur. In contrast, concentration-control coefficients, reflecting the effect of a change in an enzyme's activity on the concentration of some low molecular weight intermediate, may adopt any value. More detailed reviews of these concepts are given by Fell [29] and Cornish-Bowden[30], an explanation for molecular biologists is provided by Teusink *et al.*[42], and an example of large changes in metabolite concentrations occurring at constant flux is given by Mendes *et al*.[50]

'Top-down' MCA[31] and the related modular method[32] have some of the hierarchical features that are required for a systematic approach to the elucidation of gene function[5]. This approach aims to divide the metabolic map into a number of large units, or modules[49], through (or in) which independent fluxes or metabolite concentrations can be measured. Assignment of a novel gene as having its effect within a particular module then allows that module to be subdivided into a number of individual sub-units that can be tested, by further flux or metabolite analyses, to determine in which sub-unit the novel gene under examination has its effect. By moving down this hierarchy of analysis, a closer and closer approximation to the function of the novel gene may be obtained.

At the highest level of a top-down-MCA approach to yeast gene function, we would wish to define the whole cell as the 'system'. This means that we can define the flux-control coefficient of a particular gene product by manipulating its cellular concentration and measuring the impact of the change on growth rate, to obtain (in effect) a growth-rate-control coefficient. The experimental determination of flux control coefficients[33] demonstrates that they usually have values much closer to 0 than to 1. This means that we need a very sensitive method of measuring the change in growth rate resulting from the specific deletion of a novel yeast gene. Competition experiments between mutant and wild-type yeast have been shown to provide such a sensitive way of measuring small growth rate differences [34].

An attractive approach to competition analysis, called 'genetic footprinting', has been developed by Smith *et al*.[35]. In this approach, yeast mutants, generated by Ty1 transposon insertions, were grown in large populations under different selections using serial batch transfers to extend the period of the competition. The relative proportions of the different mutant strains in the population were monitored by PCR using a common primer complementary to the Ty sequence and a series of fluorescently labelled primers complementary to the flanking sequences of the genes whose quantitative phenotype was to be assessed. A sequencing machine equipped with GeneScan™ software was used to perform the analysis of the PCR products. This method has been used for the quantitative phenotypic analysis of yeast chromosome V[36]. Data were obtained for 261 (97%) of the predicted protein-encoding genes which showed, for almost 60%, a detectable reduction in fitness in one or more of seven different selection protocols; for the remaining genes, no phenotypic effect was found. Included in the list of phenotypes were novel, or unexpected, effects for a number of known genes. This result confirms the notion that systematic functional analysis can provide important information that will not only permit the elucidation of the function of novel genes but also provide a more integrated view of the cellular role of previously known genes.

In an approach more suited to application in top-down MCA, because measurements of growth-rate differences are made at steady-state in chemostat culures, Baganz *et al.*[37] have carried out competition experiments using specific deletion mutants. Control studies showed both that the replacement marker kanMX[10] is phenotypically neutral (in contrast to nutritional markers, such as *HIS3*[9]) and that deletion of the *HO* gene by kanMX replacement was also without phenotypic impact[37]. *HO* was chosen because it has no known role apart from mating-type switching[38], and it has been used as the site of insertion of heterologous genes in brewing yeasts without any perceptible effect on the fermentation characteristics of the organism or the quality of the product[39,40]. Whenever possible, a diploid strain homozygous for the *ho* deletion was employed, both because *HO* is inactive in diploids and because the mating of independently derived haploid transformants allows any transformation-induced genetic lesions[34] to be nullified through complementation.

In their experiments Baganz and co-workers compete test strains bearing kanMX replacements of specific ORFs with an isogenic standard strain that carries a kanMX replacement of *HO*. This allows the accurate quantitation of the impact of a specific single-gene deletion on growth rate and, moreover, the deletion mutant can subsequently be used for more specific phenotypic analyses. The proportions of mutant and standard-strain cells in the population was measured by quantitative PCR, using either densitometry or the ABI Genescan™ system. Both methods have a rather limited potential for multiplexing although, of course, DNA extracts from cultures may be aliquoted such that multiple PCR reactions may be carried out to compare the proportions of the standard strain and any number of test deletants. However, the best approach for increasing the number of mutants that may be analysed, in a single competition experiment, against the standard strain has been presented by the development of a system to generate deletion mutants that are labelled with unique oligonucleotide tags, called 'molecular bar codes'[41] that can be detected by hybridisation to a high-density oligonucleotide array[16]. This method should greatly accelerate the systematic analysis of the deletion mutants on a genome-wide scale.

The Metabolome

The second type of data required for the MCA approach is the measurement of the change in the relative concentrations of metabolites as the result of the deletion or overexpression of a gene. In contrast to the determination of flux-control coefficients, where very sensitive and discriminatory analytical tools are required, the determination of effective metabolite concentration-control coefficients requires analyses which are very comprehensive in their scope, because (for any particular novel gene) it is not known which metabolites will have their levels altered as a result of the gene's deletion or overexpression (nor even in which direction such alterations will occur[46]). Teusink *et al.*[42] proposed an analytical strategy termed FANCY (functional analysis by co-responses in yeast) which is based on the theoretical work by Rohwer *et al.*[32] and takes advantage of the fact that the number of intermediate metabolites is an order of magnitude lower than the number of genes, and that the functions of about 40% of these genes are already known.

Rohwer and co-workers have shown that some enzymes can be grouped into 'monofunctional units' in which any perturbation of the enzymes inside the unit will always produce the same coresponses outside the unit, regardless of which enzyme was affected and irrespective of the magnitude of that perturbation. This property can be exploited for functional analysis by identifying monofunctional units all over the metabolic map of yeast. For example, when two deletion mutants have the same coresponses, they will affect the same monofunctional unit. If the unit in which one of the genes causes an effect is known, that of the other can be inferred and the origin of the changes in metabolites thereby located. Such an approach requires a fast and reliable method to measure the concentrations of as many metabolites as possible to produce a 'metabolic snapshot' of each deletion mutant [42]. Oliver and co-workers[4] are developing a two-stage strategy to obtain such data. In the first phase, deletants of novel genes are grouped with those of known genes by comparing their infra-red spectra under different physiological conditions[43,44,48]. (An example of the application of FTIR spectroscopy to the analysis of mutants showing quantitative differences in respiratory activity is shown in Fig. 1.) This information then determines the type of metabolic snapshot to be taken in the second phase, using techniques such as tandem mass spectrometry[45]. The combination of the metabolic profile of a mutant together with appropriately sophisticated chemometrics[46], should allow the target site of a novel gene product (or indeed inhibitor) to be located on the metabolic map of yeast. These measurements are often best made at steady-state, which can be achieved by using chemostat cultures.

Conclusion

It is evident that a wide range of experimental approaches are being developed, for use in *S. cerevisiae,* that will allow functional genomics to build up an integrative view of the workings of a simple eukaryotic cell. This should enable a deeper understanding of more complex eukaryotes, both by the identification of orthologous genes in the different species and also by the expression of foreign coding sequences in yeast for complementation or two hybrid analyses. However, many of these techniques are sufficiently general that, once they have been tried and tested in the experimentally tractable yeast system, they should be directly applicable to the study of the functional genomics of other and higher organisms.

Acknowledgments

Work on yeast genome analysis in SGO's laboratory has been supported by the European Commission (in both the Yeast Genome Sequencing and EUROFAN Networks), the BBSRC, the Wellcome Trust, Pfizer Central Research, Applied Biosystems, Amersham International, and Zeneca. FB is grateful to Pfizer Central Research for the provision of a postgraduate research studentship. Work on the analysis of metabolism by FTIR in DBK's laboratory is supported by the BBSRC, GlaxoWellcome plc and Bruker Ltd.

References

1 Hieter, P. and Boguski, M. (1997) *Science* 278, 601-602
2 Goffeau, A. *et al.* (1996) *Science* 274, 546-567
3 Goffeau A. *et al*. (1997) *Nature* 387, 5-105
4 Oliver, S. G. (1997) *Microbiology-UK* 143, 1483-1487
5 Oliver, S. G. (1996) *Nature* 379, 597-600

6 Oliver, S. G. (1997) *Curr. Opinion Genet. & Dev.* 7, 405-409
7 Oliver, S. G. (1996) *Trends Genet.*12, 241-242
8 Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. and Cullin, C. (1993) *Nucleic Acids Res.* 21, 3329-3330
9 Baudin-Baillieu, A., Guillemet, E., Cullin, C. and Lacroute, F. (1997) *Yeast* 13, 353-356
10 Wach, A., Brachat, A., Pohlmann, R. and Philippsen, P. (1994) *Yeast* 10, 1793-1808
11 Amberg, D. C., Botstein, D. and Beasley, E. M. (1995) *Yeast* 11, 1275-1280
12 Wach, A. (1996) *Yeast* 12, 259-265
13 Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995) *Science* 38, 484-487
14 Velculescu, V. E. *et al.* (1996) *Cell* 88, 243-251
15 Kal, A. J. (1997) *Transcriptional regulation of genes encoding peroxisomal proteins in Saccharomyces cerevisiae*. Ph. D. Thesis, University of Amsterdam.
16 Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. and Davis, R. W. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93, 10614-10619
17 Lockhart, D. J. *et al.* (1996) *Nature Biotechnol.* 14, 1675-1680
18 DeRisi, J. L., Vishwanath, R. I. and Brown, P. O. (1997) *Science* 278, 680-686.
19 Hoheisel, J. D. (1997) *Trends Biotechnol.* 15, 465-469.
20 Wodicka, L., Dong, H., Mittmann, M. and Lockhart, D. J. (1997) *Nature Biotechnol.* 15, 1359-1367.
21 Wilkins, M. R. *et al.* (1996) *Bio/Technology* 14, 61-65
22 Sagliocco, F. *et al.* (1996) *Yeast* 12, 1519-1534
23 Shevchenko, A. *et al.* (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93, 14440-14445
24 Fey, S. J. *et al.* (1997) *Electrophoresis* 18, 1361- 1372
25 Bartel, P. L., Roecklein, J. A., SenGupta D. and Fields S. (1996) *Nature Genet.* 12, 72-77
26 Fromont-Racine, M., Rain, J. C. and Legrain, P. (1997) *Nat. Genet.* 16, 277-282
27 Kacser, H. and Burns, J. A. (1973) *Symp. Soc. Exp. Biol.* 27, 65-104
28 Heinrich, R. and Rapoport, T. (1974) *Eur. J. Biochem.* 42, 89-95
29 Fell, D. A. (1996). *Understanding the Control of Metabolism*. Portland Press, London.
30 Cornish-Bowden, A. (1995) *Adv. Mol. Cell. Biol.* 11, 21-64
31 Brand, M. D. (1996) *J. Theor. Biol.* 182, 351-360
32 Rohwer, J. M., Schuster, S. and Westerhoff, H. V. (1996) *J. Theor. Biol.* 179, 213-228
33 Niederberger, P., Prasad, R., Miozzari, G. and Kacser, H. (1992) *Biochem. J.* 287, 473-479
34 Danhash, N., Gardner, D. C. J. and Oliver, S. G. (1991) *Bio/Technology* 9, 179-182
35 Smith, V., Botstein, D. and Brown, P. O. (1995) *Proc. Natl. Acad. Sci.USA* 92, 6479-6483
36 Smith, V., Chou, K. N., Lashkari, D., Botstein, D. and Brown, P. O. (1996) *Science* 274, 2069-2074
37 Baganz, F., Hayes, A., Marren, D., Gardner, D. C. J. and Oliver, S.G. (1997) *Yeast* 13, 1563-1573.

38 Haber, J. E. (1992) *Trends Genet*. 8, 446-452
39 Hammond, J. R. M., Lancashire, W. D., Meaden, P. G., Oliver, S. G. and Smith, N. A. (1994) MAFF (Ministry of Agriculture, Fisheries and Food; UK) Report 07/63M
40 Yocum, R. (1986) *Proc. Bio Expo* 86, 17
41 Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M., and Davis, R. W. (1996) *Nat. Genet.* 14, 450-456
42 Teusink, B., Baganz, F., Westerhoff, H. V. and Oliver, S. G. (1998) in *Meth. in Microbiol.* **26**, 297-336.
43 Goodacre, R., Timmins, E. M., Rooney, P. J., Rowland, J. J. and Kell, D. B. (1996) *FEMS Microbiol. Lett.* 140, 233-239
44 Naumann, D., Schultz, C. and Helm, D. (1996) in *Infrared Spectroscopy of Biomolecules* (Mantsch, H. H. and Chapmann, D., eds), pp. 279-310, Wiley-Liss.
45 Rashed, M. S., Oznand, P. T., Bucknall, M. P. and Little, D. (1995) *Pediatr. Res.* 38, 324-331
46 Kell, D. B. & Westerhoff, H. V. (1986) FEMS Microbiol. Rev. 39, 305-320.
47 Goodacre, R. & Kell, D. B. (1996) Curr. Opinion Biotechnol. 7, 20-28.
48 Winson, M. K., Goodacre, R., Timmins, É. M., Jones, A., Alsberg, B. K., Woodward, A. M., Rowland, J. J. & Kell, D. B. (1997) Anal. Chim. Acta 348, 273-282.
49 Kell, D. B., van Dam, K. & Westerhoff, H. V. (1989) Symp. Soc. Gen. Microbiol. 44, 61-93.
50 Mendes, P., Kell, D. B. & Westerhoff, H. V. (1996) BBA 1289, 175-186.

Table 1: **Current Status of the EUROFAN1 Deletion Analysis**

Novel ORFs deleted	796	
Essential ORFs identified	99	(12%)
Deletion cassettes cloned	667	
Cognate genes cloned	607	

Figure 1. Preliminary analysis of the metabolome via FTIR

Colonies of *Saccharomyces cerevisiae* cells grown with glucose or glycerol as main carbon source were analysed by diffuse reflectance absorbance FT-IR (four replicates) in the mid-IR range (4000-600 cm^{-1}). Principal Components Analysis (PCA) was performed on the spectral data to reduce dimensionality whilst preserving most of the variance. Discriminant function analysis (DFA) was then used to separate the samples into groups of replicates.

Figure 1A. An illustration of some of the raw spectra (7 glucose-grown strains)

FTIR Spectra of Glucose-grown *Saccharomyces cerevisiae*

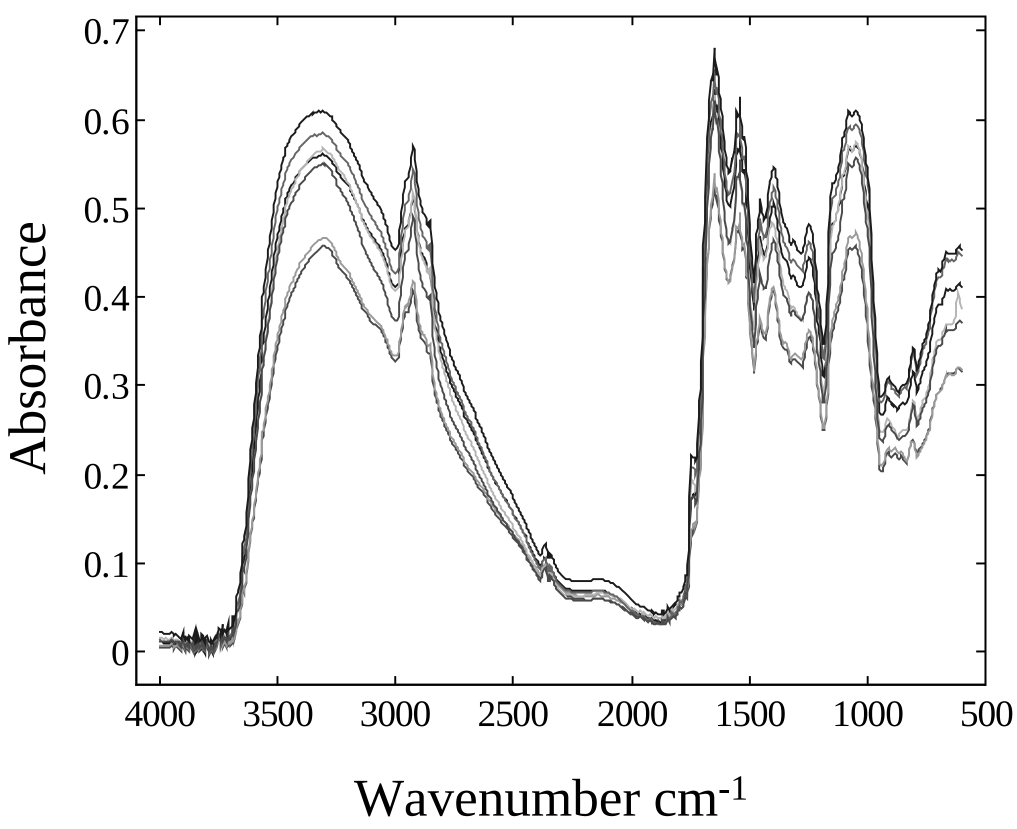
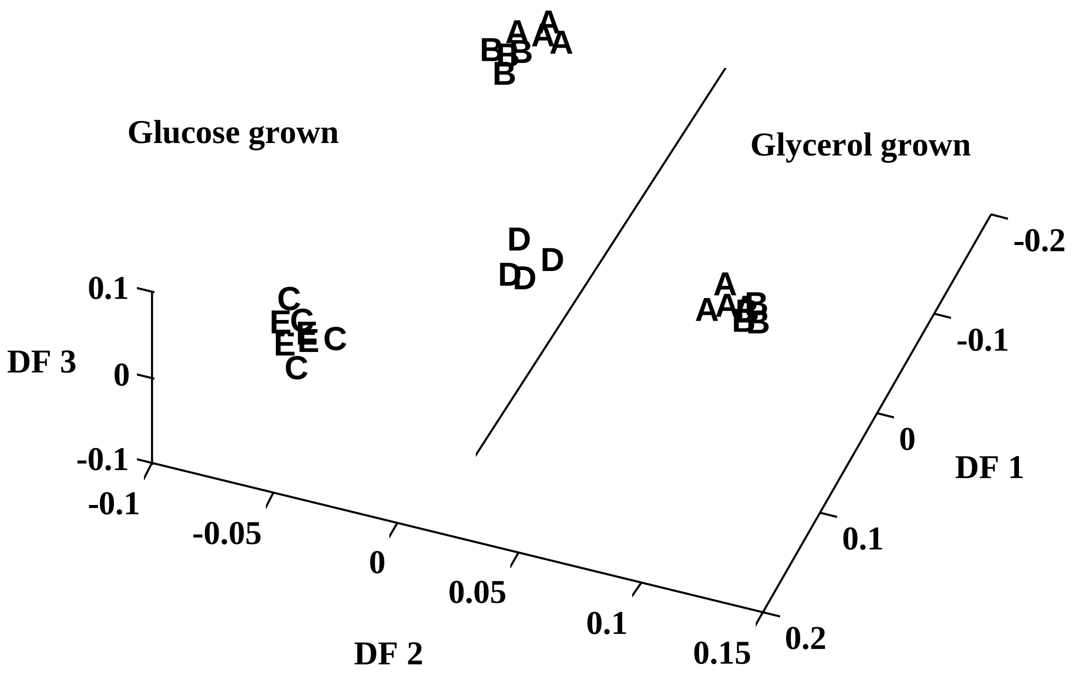


Figure 1B. Result of PCA and DFA showing clustering of strains according to growth conditions and mutant type



A. FY23 wild-type; B. FY23 Δ*ho*; C. FY23 *pet191*; D. FY23 *cox5A*; E. FY23 ρ^-

The 100% respiratory deficiency mutants FY23 *pet191::kanMX4* (C) and FY23ρ^- (E) cluster together and are separated from the partial respiratory-deficient mutant FY23*cox5a::kanMX4* (D), and *ho*(B) mutant and the FY23 wild-type (A). The strains grown on glycerol (A, B) group together but separately from glucose-grown cells.