



The yin and yang of yeast: biodiversity research and systems biology as complementary forces driving innovation in biotechnology

Ian N. Roberts, Stephen G. Oliver

► To cite this version:

Ian N. Roberts, Stephen G. Oliver. The yin and yang of yeast: biodiversity research and systems biology as complementary forces driving innovation in biotechnology. *Biotechnology Letters*, 2010, 33 (3), pp.477-487. 10.1007/s10529-010-0482-7 . hal-00647404

HAL Id: hal-00647404

<https://hal.science/hal-00647404>

Submitted on 2 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Yin and Yang of Yeast: Biodiversity Research and Systems Biology as Complementary Forces Driving Innovation in Biotechnology

Ian N Roberts^{1*} and Stephen G Oliver²

¹National Collection of Yeast Cultures, Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK

²Cambridge Systems Biology Centre and Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK

*corresponding author: ian.roberts@bbsrc.ac.uk

Introduction

The aim of this article is to review how yeast has contributed to contemporary biotechnology and to seek underlying principles relevant to its future exploitation for human benefit. Recent advances in systems biology combined with new knowledge of genome diversity promise to make yeast the eukaryotic workhorse of choice for production of everything from probiotics and pharmaceuticals to fuels and chemicals. The ability to engineer new capabilities through introduction of controlled diversity based on a complete understanding of genome complexity and metabolic flux is key. Here, we briefly summarise the history that has led to these apparently simple organisms being employed in such a broad range of commercial applications. Subsequently, we discuss the likely consequences of current yeast research for the future of biotechnological innovation.

Yeast at the forefront of biotechnology

- Mankind's earliest biotechnology

The ancient Egyptians are usually credited with the first proven use of yeast, largely as a result of the discovery of 4,000 year old hieroglyphics depicting clearly recognisable brewing and baking procedures. However, it is likely that mankind's earliest use pre-dates even the origins of agriculture. Virtually all human societies are believed to have utilised yeasts in some way or other for a variety of food and beverage fermentations. Nevertheless, it was not until the work of Louis Pasteur and others in the 1800s that the true living nature of these predominantly single-celled fungi was unambiguously defined (Barnett 2000). Subsequent rapid progress in microbiology (including the development of the first vaccines) and in biochemistry (the word "enzyme" itself is derived from the Greek words meaning "in yeast") led to the emergence of yeast as a leading model organism in biological research (Barnett 2003).

- First pure cultures and strain collections

The development of sterile technique and the adoption of pure yeast cultures in commercial brewing during the early 1900s was perhaps the most important innovation as regards the use of yeast in modern biotechnology. Recognition of variation in the growth and fermentation characteristics of different pure cultures provided the opportunity to launch targeted strain improvement programmes. The first collections of yeasts with defined characteristics were assembled by Hansen around the turn of the twentieth century (Barnett and Lichtenthaler 2001). Public domain culture collections effectively came into being with

the publication of the first catalogue of microbial strains by Král in 1900 (Uruburu 2003). Today, tens of thousands of different strains are available from culture collections worldwide.

For example, the first page of a Google search for 'yeast culture collection' in August 2010 yielded direct links to the National Collection of Yeast Cultures (UK), the Phaff Yeast Culture Collection (USA), the American Type Culture Collection (USA), the VTT culture collection (Finland) and the World Data Centre for Microorganisms (Japan). The last site links to 568 public domain culture collections in 68 countries, most notably the CBS (Netherlands) which maintains Hansen's original yeast collections alongside taxonomic type strains of more than 1,000 other yeast species described to date.

The holdings of the public domain culture collections are only the tip of the iceberg. Thousands more proprietary strains are held by private organisations in the commercial biotechnology sector. Nevertheless, these public collections are significant: they provide back-up resources to safeguard commercially important genetic variation against market failures; they offer state-of-the-art identification facilities to meet an ever-increasing demand for precision characterisation of yeast strains of biotechnological interest; and they are invaluable in the academic sphere as a source of expertise and fully-authenticated material for biological research (Boundy-Mills 2010). In short, they provide the capability for the biotechnological advances described below and, most importantly of all, guarantee the reproducibility of the scientific experimentation on which these advances are based.

- Pioneering genetic transformation work

Despite considerable progress in yeast genetics during the early 1900s to 1960s, the next most important biotechnological innovation could be considered to be the achievement of genetic transformation in the 1970s (Beggs 1978; Hinnen et al 1978). Although the use of genetically manipulated organisms in food biotechnology remains politically sensitive in some parts of the world, the use of patented genetically modified yeast to meet contemporary societal demands, such as producing pharmaceuticals and carbon neutral 'second generation' biofuels, appears to have achieved widespread acceptance. Furthermore, the application of 'reverse genetics' techniques from the 1980s onwards has rapidly accelerated fundamental understanding of basic yeast biology. The armoury of tools and resources available to the present day yeast researcher or biotechnologist is unrivalled by any other eukaryotic organism.

- First eukaryotic 'omics and establishment of large-scale databases

The small size of the *Saccharomyces cerevisiae* yeast genome relative to other eukaryotes, coupled with a century-long history as a model organism in genetics (alongside its considerable commercial and medical importance), led to its selection for the first eukaryotic whole genome sequencing project. This EU-funded project began in the early 1990s and effectively laid the scientific foundations for the human genome project and many other subsequent large-scale sequencing projects. The completion of the EU *Saccharomyces cerevisiae* genome project (Goffeau et al 1996) and its comprehensive bioinformatics analysis (Mewes et al 1997) created a need for a genome database for this yeast. There are two comprehensive *S. cerevisiae* yeast genome databases – SGD (Engel et al 2010) and CYGD (Mewes et al 2008). They take complementary approaches to the annotation of gene function and can usefully be employed in conjunction. No other biotechnologically important

yeast has a comparable database, but Table 1 provides links to genome databases for other yeast species as well as to more specialised databases that relate exclusively to *S. cerevisiae*.

The EU yeast genome project was followed by the construction of sets of gene deletion mutants for each of the approximately 6,000 genes discovered (Scherens and Goffeau 2004) and a new era of functional genomics. Since then sequencing costs have fallen dramatically. Genome re-sequencing projects (e.g. Wei et al 2007, Doniger et al 2008, Bornemann et al 2008, Dimitrov et al 2009, Liti et al 2009, Schacherer et al 2009, Dowell et al 2010), new robotic techniques and the construction of mathematical and logical models (e.g. King et al 2009) have led to the generation of a rapidly increasing number of sequence databanks, genetic resources and computational tools, to the point that the yeast-associated scientific infrastructure, including culture collections and data centres, has expanded beyond recognition. The NCBI Entrez Genome Project database now lists 19 separate genome sequencing projects for *Saccharomyces cerevisiae* alone. Entries include several commercially important wine strains in addition to the laboratory and environmental isolates that were the focus of the earlier studies. Vast volumes of new information relevant to future biotechnological innovation continue to be produced. Two examples are given below.

- Leading computational models of cell function

The history of biotechnological innovation shows a clear dependence upon advances in understanding of fundamental biology. Early exploitation of strain variation depended upon isolation of pure cultures and recognition of their biological nature. Later innovations depended upon modification of biochemical pathways based on scientifically-established genetic principles. The next major advances will probably derive from the latest 'systems' level understanding of cellular organisation emerging from computational genomics. This is one motivation behind current research to develop mathematical models of the entire biochemical repertoire of a single cell. The latest models of gene expression, biochemical pathways and metabolic flux through cellular networks are leading to the ability to apply engineering design principles to achieve high level control, thus expediting strain improvement. Yeast is again leading the way with the most accurate computational model of a eukaryotic cell metabolism achieved to date (Herrgård et al 2008). Complementary progress in understanding yeast biodiversity, genome variation and genome dynamics is also being achieved at a similar rate through the application of 'next generation' genomics (see below).

- Most complete knowledge of genome variation

Rapid technological advances in DNA sequencing and associated informatics have led to the term 'next generation' genomics being coined to describe a huge leap in data gathering capacity. The pace of advance in terms of cost per base and speed of genome sequencing, assembly and comparative analysis shows no sign of diminishing. Computational comparative analysis and mathematical descriptions of genome variation and dynamics are providing a deep knowledge base and a broad sampling of yeast diversity to help guide biotechnological exploitation. For example, the *Saccharomyces* Genome Re-sequencing Project (Liti et al 2009) has sampled far greater biological diversity than for any other eukaryote, including many strains isolated from "wild" environments as well as those already used across the world in a variety of commercial processes. Yeast genomes are easily and

cheaply sequenced, and the genome analysis techniques pioneered with yeasts are of widespread applicability in health and agriculture. Further advances in genomics and systems biology mean that it will become facile to engineer a much wider range of yeast species. The genomic diversity available for future exploitation is discussed below.

Yeast's extraordinary biological diversity

- Fungal origins and phylogeny

Yeasts are classified biologically as fungi. Kurtzman and Fell (1998) define yeasts as "...fungi with vegetative states that predominantly reproduce by budding or fission, resulting in growth that is comprised mainly of single cells". Budding yeasts belonging to the ascomycetes group are monophyletic and comprise about 1,000 different species. These species are surprisingly diverse phylogenetically (Suh et al 2006) and have recently been assigned to 14 different lineages on the basis of multigene sequence analysis (Figure 1). The remainder of yeast species fall into more distantly-related fungal groupings, most notably the basidiomycetes (Scorzetti et al 2002), which includes an expanding list of yeast species with properties of biotechnological interest (e.g. Zhang et al 2009). These novel species add significantly to the overall knowledge of yeast phylogenetic diversity and hence to the 'gene pool' available for exploitation. It is also clear that many more yeast species remain to be discovered. Beetle guts are proving a particularly rich source of hitherto undescribed species, potentially encompassing a range of novel biocatalytic activities of value to second-generation biofuel production (Suh et al 2005).

- Evolutionary breadth

In addition to their phylogenetic diversity, whole-genome sequencing of representatives of different genera (Figure 1) has shown that yeasts span a huge evolutionary range. Distributions of amino acid identity between pairs of orthologous proteins indicate that *S. cerevisiae* is as diverged from *Yarrowia lipolytica* as *Homo sapiens* is from the sea squirt, *Ciona intestinalis* (Dujon 2006). The first organisms comparable to modern yeasts are thought to have evolved over 100 million years ago, coinciding with the evolutionary radiation of flowering plants (Wolfe and Shields 1997) and around the time that the whole-genome duplication observed in *S. cerevisiae* is thought to have occurred (Figure 1). However, yeast-like organisms were probably active in the environment long before this time. From this, we conclude that an extraordinary metabolic potential is available for exploitation and the vast majority of this potential has yet to be tapped.

- Sub-species diversity

We are also only just beginning to grasp the equally extraordinary variation that exists within a single yeast species. Large-scale genome sequencing of over seventy *S. cerevisiae* isolates representative of global population diversity revealed that there are at least four separate geographical lineages worldwide and considerable variation between and within each lineage (Liti et al 2009). Detailed examination of strains used commercially also reveals extensive cross hybridisation (Dunn and Sherlock 2008), genomic mosaicism (Liti et al 2009) and sub-species diversity in gene copy number variation, all arising from artificial selection for novel traits (Stambuk et al 2009). Similar variation is to be expected within all the other 1,000 or so yeast species described to date and the many more that remain to be discovered.

- Individual cell stochasticity

Even within a yeast culture grown from a single colony, it is becoming clear that considerable biological variation is present. The genome sequence and environmental conditions may remain constant but noise in cellular transcription or translation processes (stochasticity) gives rise to phenotypic individuality. This individuality is manifested in phenomena such as the existence of sub-populations with elevated resistance to environmental challenges e.g. ability to survive in the presence of antimicrobial chemicals (Avery 2006).

While a systems-level understanding of exactly how all the above variation arises has yet to be achieved, it is already clear that yeasts are an exceptionally diverse biological group with high “evolvability”. This helps explain the numerous biotechnological applications which already utilise yeast. It is not our intention here to provide a comprehensive list of such applications, these can be found summarised throughout the yeast-related literature (e.g. Wolf 1996). Rather, we will select a subset of such applications and discuss the innovative potency of combining yeast systems biology with yeast genome diversity.

Future prospects for yeast biotechnology

- Probiotics

Yeast growth in the human gut is generally thought to be an indicator of ill health. However, there is one yeast strain that is currently used widely as a therapeutic agent (Zanello et al 2009). The probiotic yeast *Saccharomyces boulardii* is taxonomically indistinguishable from strains of *S. cerevisiae*, yet is able to suppress intestinal inflammation caused by a broad range of enteric pathogens. *S. boulardii* has been shown to be particularly useful in the treatment of antibiotic-induced diarrhoea and *Clostridium difficile* infection. The yeast strain appears to act by modulating host signalling pathways. Inhibition of interleukin-8 production, which is otherwise induced by *C. difficile* toxin, leads to reduced intestinal inflammatory responses (Zanello et al 2009). Since *S. boulardii* is not typically an enteric organism, understanding its special characteristics and its effects on survival of gut bacteria will provide a useful tool to probe probiotic functions of yeasts in the gut.

Recent metagenomic analyses of gut microflora have suggested that in the region of 0.1% of sequences sampled are of eukaryotic origin (Qin et al 2010). However, the results of Scanlan and Marchesi (2008) indicate that this may be an underestimate. Culturing and DNA sequencing techniques appear to give different results in terms of the eukaryotic microbial species identified. There may therefore be a risk of sampling errors due to the use of DNA extraction techniques which are efficient for bacteria but sub-optimal for fungi. It is also an open question how the reduction in complexity of prokaryotic flora associated with ageing is matched by changes in the eukaryotic flora. However, healthy ageing may well be contingent upon such changes. Furthermore, the recent European Food Safety Authority (EFSA 2009) ruling that health claims for probiotics must be supported by improved characterisation will drive further research into characterisation of yeast-derived bioactive compounds in the gut, with clear opportunities for further biotechnological innovation. Indeed, there is already evidence of strain variation with respect to immunological effects in the mammalian gut (Foligne et al 2010).

- Pharmaceuticals

Yeast strains have been used for many years as cell factories e.g. for insulin production (Pscheidt and Glieder 2008). Innovation in pharmaceutical interventions has been achieved through improved understanding of membrane transporters based on studies in yeast (Kim and Fay 2009). Yeasts have also become favoured vehicles for the production of recombinant vaccines. Hepatitis B vaccine has been produced in *S. cerevisiae* for some time (McAleer et al 1984) and two companies now use this yeast to produce a vaccine against human papilloma virus (HPV) in order to protect against cervical cancer (Romanos et al 1995; Neeper et al 1996; Hofmann et al 1996; Bryan 2007). *S. cerevisiae* is able to assemble capsid subunit proteins into virus-like proteins; however, for the HPV vaccine, it has still proved necessary to re-fold the protein and assemble the particle in vitro. An alternative strategy, of using yeast as a vehicle to deliver vaccines to defined portions of the gut by genetically controlled cell lysis, may be worth pursuing (Omara et al 2010). Vaccines apart, *Pichia pastoris* has become the favoured yeast as a vehicle for the production of therapeutic proteins, both due to its higher protein secretion capacity and to the very high cell densities that it achieves in industrial fermenters (Cereghino and Cregg 2000; Li et al 2007; Mattanovich et al 2009; Sohn et al 2010).

Yeast has also been subjected to extensive pathway engineering in order to allow it to produce therapeutic agents that defy chemical synthesis and are often difficult to isolate in sufficient quantity from their natural producers, examples include human hydrocortisone (Szczepara et al 2003) and the antimalarial plant product, artemisinin (Ro et al 2006). Problems of inappropriate glycosylation of secreted proteins have been largely overcome (De Pourcq et al 2010). In the future, it can be expected that pathway engineering will be extended to generate products that are either hybrids of compounds produced by other species or to synthesise completely novel products.

- Fuels

'First generation' biofuel production suffered a major unintended drawback as competition for raw materials resulted in raised food prices in many parts of the world. 'Second generation' bioethanol production seeks to use non-food components of crop plants e.g. wheat straw. However, this aim is complicated by the need to break down complex lignocellulosic plant cell wall material. It is questionable whether cost-effective production of fermentable sugars comparable to the substrates used in first-generation processes can be achieved in the short term. Novel processes will likely depend upon long-term breeding of new crop varieties with modified cell walls (see, for example, Bouton, 2007). However, in the interim, yeasts capable of fermenting a wider range of sugars and capable of overcoming fermentation inhibitors present in crude plant cell wall extracts are actively being sought.

A recent example of how yeast's metabolism might be engineered to overcome a well-known inhibitor of fermentation present in lignocellulosic substrates (acetic acid) is given by Guadalupe Medina et al (2010). Their engineered strain also has the highly desirable commercial property of not producing glycerol as a by-product of fermentation. While many other yeast strains have been successfully engineered specifically for commercial biofuel production (e.g. Alper et al 2006), a much broader range of activities is still required. Efficient fermentation of substrates containing glucose, arabinose and xylose in varying ratios and with considerable variation between batches may call for mixed cultures of different yeast

species. Bioaugmentation with bacterial species may well have a role in order to break down any oligosaccharides and fermentation inhibitors still remaining after chemical or physical pre-treatment of the raw material. Furthermore, combustion enhancers, such as specific volatiles present in small quantities following fermentation, could be provided by different yeast species as a co-product of mixed fermentations. These could aid the development of 'next generation' engines capable of running on a 'weaker mix' of bioethanol. Such progress will depend on innovative, robotised selection of genomes, strains and strain mixtures optimised for fermentation of cruder substrates and resistant to batch variations.

Efficient yeast bioprocesses for biofuel production will require that strain development takes a holistic view of how to channel carbon flux toward the desired product. When carbon sources are used which are not natural substrates for yeast, there will often be problems such as cofactor recycle or redox balance to contend with. The classic example is the problem of engineering *S. cerevisiae* to use pentose sugars from lignocellulose hydrolisates. Here, it is necessary to engineer the organism for optimal uptake of pentoses by recruiting a heterologous transporter (Saloheimo et al, 2007), as well as solve the re-cycling problems due to the different cofactor requirements of the enzymes recruited to perform the first two steps in xylose catabolism (Verho et al 2003). Finally, classic carbon catabolite repression is not observed when cells are grown on xylose (Salusjärvi et al 2008). All of this means that it is imperative to take a systems biology approach and use a genome-scale metabolic model to direct the metabolic engineering of the production strain (Bro et al 2006).

- Minimisation and humanisation

As more and more yeast genomes are sequenced, it is becoming possible to define with increasing accuracy what may be described as the 'minimal yeast genome'. For example, reduction in the copy number of ribosomal DNA from around 150 copies to less than 40 copies appears to have no effect on cell viability (Kobayashi et al 1998), although it has recently been shown to impact on genome integrity (Ide et al 2010). Similarly, not all strains have identical gene complements so strain comparisons enable a 'core yeast genome' to be defined. Such information provides the raw material to engineer a yeast cell with a minimal gene complement. With a minimal genome in place and a systems-level understanding of its 'wiring diagram' and 'design features' it is possible to envisage simple cellular network re-programming e.g. by engineering modified histone proteins enabling low temperature growth (c.f. Kumar and Wigge 2010) or by introducing phosphorylation sites to modify signalling cascades controlling responses to environmental fluctuations, in order to expand the range of conditions in which yeasts may be used. While the complexity of even a simple eukaryote will mean that unexpected emergent properties will still arise, such cells offer the potential for far more targeted and controllable approaches to achieving novel biotechnological applications.

Furthermore, yeast has still to fulfil its full potential as a medical research tool reducing the need for animal experimentation. Yeast could be used as a 'living test-rig' to identify the parameters of human metabolism and determine how these change in disease states. To achieve this, yeast components of specific pathways and systems (mainly enzymes) would be replaced with the coding sequences for their human equivalents under the control of the promoter of the yeast ortholog. This would allow the kinetic parameters of human enzymes to be inferred from changes in the system properties when the system is modulated by, for

example, toxins. These observations would be used as the basis for predictions about the behaviour of the human pathway, which could then be tested in human cell lines.

Summary

We have attempted to define ways in which systems biology and biodiversity research complement each other and can be combined to drive innovation in biotechnology. Ever since Hansen's early contributions, strategic development of culture collections has underpinned biotechnological exploitation of yeast. As noted previously (Barnett and Lichtenthaler 2001), although Hansen made no obvious scientific advances compared to, for example, Fischer and Buchner of the same era, the latter scientists would have made little progress without the resources provided and the relatively new techniques of microscopy and microbiology that were emerging at the same time. In the post-genome era, culture collections have taken on an expanded role as genetic resource centres. Hitherto untapped biodiversity and screening facilities are being made readily available alongside expertise in genomic analysis. Appropriately formatted collections of sub-species diversity, combined with revolutionary advances in next-generation DNA sequencing, systems and synthetic biology, and the rapidly evolving data resources available as web services, promise to drive a multitude of new applications in fundamental research, biotechnology, sustainable agriculture, and human healthcare.

Acknowledgements

The authors would like to acknowledge financial support from the BBSRC (INR and SGO) and the European Commission through the UNICELLSYS Project (SGO). We also thank Cleve Kurtzman and Ken Wolfe for advice and Tom Mayle of the Cambridge University Graphics Department for preparation of Figure 1.

References

- Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production *Science* 314: 1565-8
- Avery SV (2006) Microbial cell individuality and the underlying sources of heterogeneity *Nat. Rev. Microbiol.* 4: 577-87
- Barnett JA (2000) A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850-1880 *Yeast* 16: 755-71
- Barnett JA (2003) Beginnings of microbiology and biochemistry: the contribution of yeast research *Microbiology* 149: 557-67
- Barnett JA, Lichtenthaler FW (2001) A history of research on yeasts 3: Emil Fischer, Eduard Buchner and their contemporaries, 1880-1900 *Yeast* 18: 363-88
- Beggs JD (1978) Transformation of yeast by a replicating hybrid plasmid *Nature* 275:104-9
- Borneman AR, Forgan AH, Pretorius IS, Chambers PJ (2008) Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain *FEMS Yeast Res.* 8, 1185-95
- Boundy-Mills K (2010) Yeast culture collections of the world: More than a source of cultures *SIM News* (In press)

- Bouton JH (2007) Molecular breeding of switchgrass for use as a biofuel crop *Curr Opin Genet Dev* 17: 553–558
- Bro C, Regenber B, Förster J, Nielsen J (2006) In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng* 8:102–111
- Bryan JT (2007) Developing an HPV vaccine to prevent cervical cancer and genital warts *Vaccine* 25: 3001-3006
- Cereghino, JL and Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris* *FEMS Microbiol. Revs.* 24: 45-66
- De Pourcq K, De Schutter K, Callewaert N (2010) Engineering of glycosylation in yeast and other fungi: current state and perspectives *Appl Microbiol Biotechnol.* 87: 1617-31
- Dimitrov LN, Brem RB, Kruglyak L, Gottschling DE (2009) Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of *Saccharomyces cerevisiae* S288C strains *Genetics* 183, 365-83
- Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang SP, Fay JC (2008) A catalog of neutral and deleterious polymorphism in yeast *PLoS Genet* 4, e1000183
- Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, Bernstein DA, Rolfe PA, Heisler LE, Chin B, Nislow C, Giaever G, Phillips PC, Fink GR, Gifford DK and Boone C (2010) Genotype to phenotype: a complex problem *Science* 328, 469
- Dujon B (2006) Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution *Trends Genet.* 22: 375-87
- Dunn B, Sherlock G (2008) Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus* *Genome Res.* 18: 1610-23
- EFSA (2009) EFSA delivers its first series of opinions on ‘general function’ health claims http://www.efsa.europa.eu/EFSA/efsa_locale-1178620753812_1211902914361.htm
- Engel SR, Balakrishnan R, Binkley G, Christie KR, Costanzo MC, Dwight SS, Fisk DG, Hirschman JE, Hitz BC, Hong EL, Krieger CJ, Livstone MS, Miyasato SR, Nash R, Oughtred R, Park J, Skrzypek MS, Weng S, Wong ED, Dolinski K, Botstein D, Cherry JM (2010) *Saccharomyces* Genome Database provides mutant phenotype data *Nucleic Acids Res.* 38: D433-6
- Foligné B, Dewulf J, Vandekerckove P, Pignède G, Pot B (2010) Probiotic yeasts: anti-inflammatory potential of various non-pathogenic strains in experimental colitis in mice *World J Gastroenterol.* 16:2134-45
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes *Science* 274: 563-7
- Guadalupe Medina V, Almering MJ, van Maris AJ, Pronk JT (2010) Elimination of glycerol production in anaerobic cultures of a *Saccharomyces cerevisiae* strain engineered to use acetic acid as an electron acceptor *Appl Environ Microbiol.* 76:190-5

Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasić I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttilä M, Klipp E, Palsson BØ, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology *Nat Biotechnol.* 26: 1155-60

Hinnen A, Hicks JB, Fink GR (1978) Transformation of yeast *Proc. Natl. Acad. Sci. USA* 75:1929-33

Hofmann KJ, Neeper MP, Markus HZ, et al. (1996) Sequence conservation within the major capsid protein of human papillomavirus (HPV) type 18 and formation of HPV-18 virus-like particles in *Saccharomyces cerevisiae* *J.Gen. Virol.* 77: 465-468

Ide S, Miyazaki T, Maki H, Kobayashi T (2010) Abundance of ribosomal RNA gene copies maintains genome integrity *Science* 327: 693-69

Kim HS, Fay JC (2009) A combined-cross analysis reveals genes with drug-specific and background-dependent effects on drug sensitivity in *Saccharomyces cerevisiae* *Genetics* 183: 1141-51

King RD, Rowland J, Oliver SG, Young M, Aubrey W, Byrne E, Liakata M, Markham M, Pir P, Soldatova LN, Sparkes A, Whelan KE, Clare A (2009) The automation of science *Science* 324: 85-9

Kobayashi T, Heck DJ, Nomura M, Horiuchi T (1998) Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I *Genes Dev.* 12: 3821–3830

Kumar SV, Wigge PA (2010) H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in *Arabidopsis* *Cell* 140: 136-147

Kurtzman CP (2003) Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the *Saccharomycetaceae*, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygorhizula* *FEMS Yeast Res.* 4: 233-45

Kurtzman CP and Fell JW (1998) *The Yeasts - A Taxonomic Study*, Fourth Edition p3

Li P, Anumanthan, A, Gao XG, Ilangoan K, Suzara VV, Duezguenes N, RenugopalakrishnanV (2007) Expression of recombinant proteins in *Pichia pastoris* *Applied Biochemistry and Biotechnology* 142: 105-24

Liti G., Carter D. M., Moses A. M., Warringer J., Parts L., James S. A., Davey R. P., Roberts I. N., Burt A., Koufopanou V., Tsai I. J., Bergman C. M., Bensasson D., O'Kelly M. J. T., van Oudenaarden A., Barton D. B., Bailes E., Jones M., Durbin R., Louis E. J. (2009) Population genomics of domestic and wild yeasts *Nature* 458: 337-341

Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B (2009) Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris* Microbial Cell Factories 8:29

McAleer WJ, Buynak EB, Maigetter RZ, Wampler DE, Miller WJ, Hilleman MR (1984) Human hepatitis B vaccine from recombinant yeast Nature 307: 178-180

Mewes HW, Albermann K, Bähr M, Frishman D, Gleissner D, Hani J, Heumann K, Kleine K, Maierl A, Oliver SG, Pfeiffer F, Zollner A (1997) Overview of the yeast genome Nature 387 (Supp.): 7-65.

Mewes HW, Dietmann S, Frishman D, Gregory R, Mannhaupt G, Mayer KF, Münsterkötter M, Ruepp A, Spannagl M, Stümpflen V, Rattei T (2008) MIPS: analysis and annotation of genome information in 2007 Nucleic Acids Res. 36: D196-201

Neeper MP, Hofmann KJ, Jansen KU (1996) Expression of the major capsid protein of human papillomavirus type 11 in *Saccharomyces cerevisiae* Gene 180: 1-6

Omara W, Rash BM, Hayes A, Wickham MSJ, Oliver SG, Stateva LI (2010) Conditional cell wall mutants of *Saccharomyces cerevisiae* as delivery vehicles for therapeutic agents in vivo in the GI tract. J. Biotechnol. 147: 136-143.

Pscheidt B, Glieder A (2009) Yeast cell factories for fine chemical and API production Microb Cell Fact. 7: 25

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J; MetaHIT Consortium, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing Nature 464: 59-65

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MC, Withers ST, Shiba Y, Sarpong R, Keasling JD (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast Nature 440: 940-3

Romanos MA, Hughes FJ, Comerford SA, Scorer CA (1995) Production of a phosphorylated GST::HPV-6 E7 fusion protein using a yeast expression vector and glutathione S-transferase fusions Gene 152: 137-138

Saloheimo A, Rauta J, Stasyk OV, Sibirny AA, Penttilä M, Ruohonen L (2007) Xylose transport studies with xylose-utilizing *Saccharomyces cerevisiae* strains expressing heterologous and homologous permeases. Appl Microbiol Biotechnol 74: 1041–1052

Salusjärvi L, Kankainen M, Soliymani R, Pitkänen J-P, Penttilä M, Ruohonen L (2008) Regulation of xylose metabolism in recombinant *Saccharomyces cerevisiae* Microb Cell Factories 7:18 (16 pages)

Scanlan PD, Marchesi JR (2008) Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces ISME J. 2:1183-93

Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH (2007) Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication Proc. Natl. Acad. Sci. USA 104: 8397–8402

Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae* Nature 458, 342-5

Scherens B, Goffeau A (2004) The uses of genome-wide yeast mutant collections Genome Biol. 5: 229

Scorzetti G, Fell JW, Fonseca A, Statzell-Tallman A. (2002) Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions FEMS Yeast Res. 2: 495-517

Sohn SB, Graf AB, Kim TY, Gasser B, Maurer M, Ferrer P, Mattanovich D, Lee SY (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production Biotechnol J. 5: 705-15

Stambuk BU, Dunn B, Alves SL Jr, Duval EH, Sherlock G (2009) Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis Genome Res. 19: 2271-8

Suh SO, McHugh JV, Pollock DD, Blackwell M (2005) The beetle gut: a hyperdiverse source of novel yeasts Mycol Res. 109: 261-5

Suh SO, Blackwell M, Kurtzman CP, Lachance MA (2006) Phylogenetics of Saccharomycetales, the ascomycete yeasts Mycologia 98: 1006-17

Szczębara FM, Chandelier C, Villeret C, Masurel A, Bourrot S, Duport C, Blanchard S, Groisillier A, Testet E, Costaglioli P, Cauet G, Degryse E, Balbuena D, Winter J, Achstetter T, Spagnoli R, Pompon D, Dumas B (2003) Total biosynthesis of hydrocortisone from a simple carbon source in yeast Nature Biotechnol 21: 143-149

Uruburu F (2003) History and services of culture collections International Microbiology 6: 101-103

Verho R, Londesborough J, Penttilä M, Richard P (2003) Engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae* Appl Environ Microbiol. 69: 5892-7

Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, Cao Z, Gu Z, Bruno D, Miranda M, Nguyen M, Wilhelmy J, Komp C, Tamse R, Wang X, Jia P, Luedi P, Oefner PJ, David L, Dietrich FS, Li Y, Davis RW, Steinmetz LM (2007) Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789 Proc Natl Acad Sci U S A 104, 12825-30

Wolf K (1996) Nonconventional Yeasts in Biotechnology: A Handbook (Springer)

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome Nature 387:708-13

Zanello G, Meurens F, Berri M, Salmon H (2009) *Saccharomyces boulardii* effects on gastrointestinal diseases Curr Issues Mol Biol. 11: 47-58

Zhang X, Zhang H, Li X, Su Z, Wang J, Zhang C (2009) Isolation and characterization of *Sporobolomyces* sp. LF1 capable of degrading chlorimuron-ethyl J Environ Sci (China) 21: 1253-60

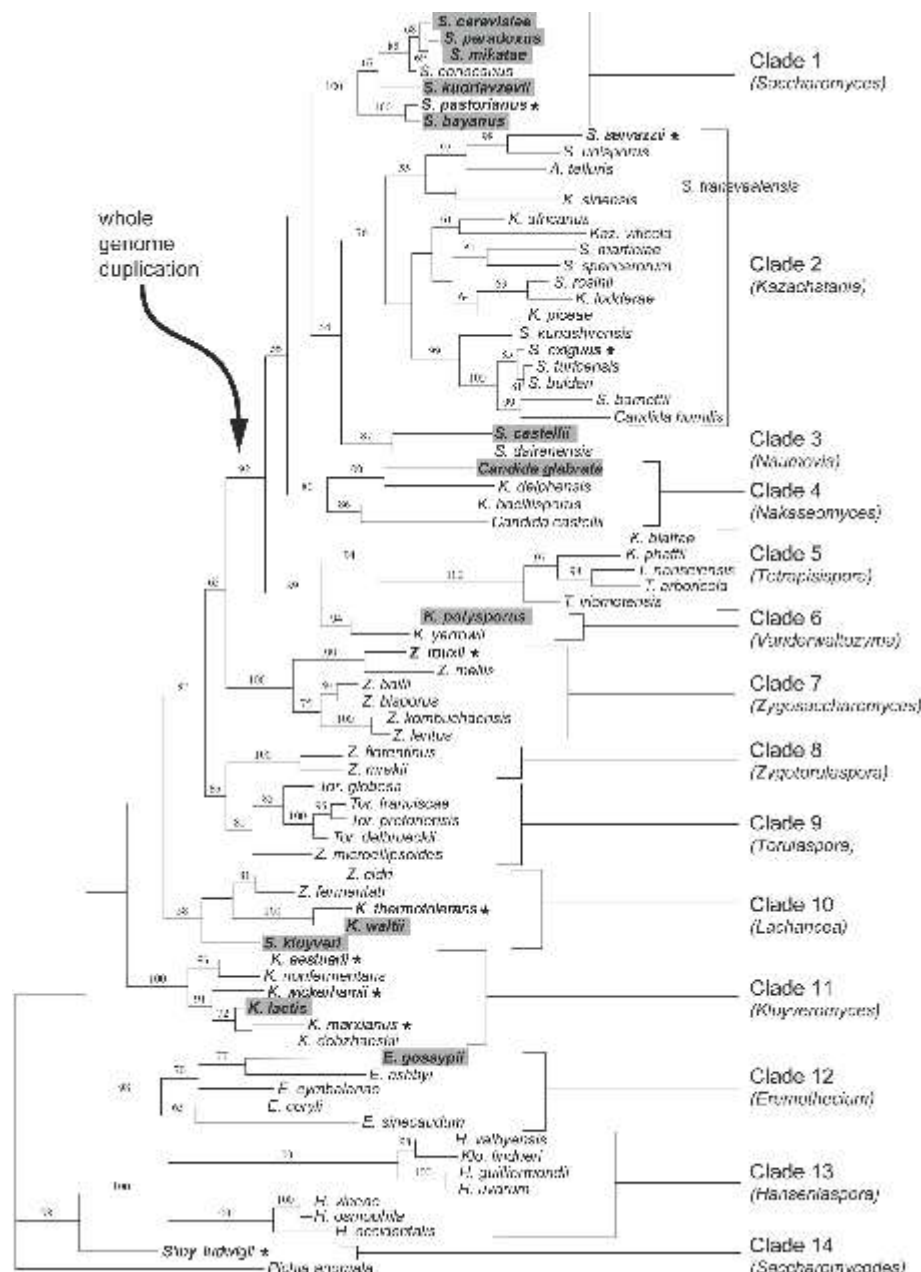


Figure 1: Yeast phylogenetic tree from Kurtzman (2003) re-drawn by Scannell et al (2007) (see Scannell et al 2007 Supplementary Information: Figure 5) to include information on species with completed genome sequences (bold type; shaded background) and inferred position of divergence due to whole genome duplication (arrow). The figure has been updated to include August 2010 information on additional completed and ongoing yeast

genome sequencing projects (bold type; asterisked). Information was taken from www.genolevures.org and <http://www.ncbi.nlm.nih.gov/nuccore> (search term: fungi[orgn] AND wgs_master[prop]). The latter search identified 31 other yeast species not represented on the tree as having undergone or currently undergoing whole genome sequencing.

Table 1 Databases and other on-line resources useful for research on yeasts of biotechnological importance

| DATABASE | Comments | URL |
|---|----------|--|
| SGD (<i>Saccharomyces</i> Genome database) | | http://www.yeastgenome.org/ |
| | | The most widely used yeast genome database. Employs GO (Gene Ontology) for functional annotation. Has a wide range of data and community resources and useful links. |
| CYGD (Comprehensive Yeast Genome Database) | | http://mips.helmholtz-muenchen.de/genre/proj/yeast/ |
| | | Employs its own system (FunCat) for functional annotation, which has a number of useful features that are complementary to GO. A particularly useful database for features relating to |

protein
sequence and
structure.

Saccharomyces Genome Resequencing Project
(SGRP) <http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html>

Contains
genome
sequence data
for *S. cerevisiae* and
S. paradoxus.

Includes
sequence data
for a number
of wild and
industrial
strains,
including wine
yeasts.

Génolevures <http://www.genolevures.org/>

Contains
genome
sequences for
>30
hemiascomycete yeasts.
Many are of
industrial
importance,
and direct
links to the
databases for
some of these
species are
given below.

Kluyveromyces lactis
genome
sequence <http://www.genolevures.org/klla.html>

The *K. lactis* genome sequence from Génolevures. Also provides some background information about the species.

Yarrowia lipolytica
genome
sequence <http://www.genolevures.org/yali.html>

The *Y. lipolytica* genome sequence from Génolevures. Also provides some background information about the species.

Pichia pastoris <http://www.pichiagenome.org/>
genome
browser

Until recently, there was no publicly accessible genome sequence for this hugely biotechnologically important yeast species. Now there are two – this from the group of organism expert Diethard Mattanovich.

Pichia pastoris <http://bioinformatics.psb.ugent.be/webtools/bogas/overview/Picpa>
genome
portal

A very well-annotated *P. pastoris* sequence from leading bioinformatician Yves van der Peer

Yeast Gene <http://wolfe.gen.tcd.ie/ygob/>
Order
Browser
(YGOB)

An easy-to-use
on-line viewer
that enables
syntenic
comparisons
between 16 yeast
species and also
a reconstruction
of their ancestral
genome.

| | | |
|---|---|---|
| EUROSCARF (European Saccharomyces cerevisiae Archive for Functional Analysis) | http://web.uni-frankfurt.de/fb15/mikro/euroscarf/index.html | Index to the <i>S. cerevisiae</i> strains and plasmids repository. Originally built for the EUROFAN Project, it contains the complete deletion collection and many more biological resources. |
| Profiling of phenotypic characteristics in yeast (PROPHECY) | http://prophecy.lundberg.gu.se/ | Quantitative phenotypic information for the complete set of <i>S. cerevisiae</i> deletion mutants. |
| NetPhosyeast | http://www.cbs.dtu.dk/services/NetPhosYeast/ | An on-line tool to predict phosphorylation sites in <i>S. cerevisiae</i> proteins. |
| BioGRID | http://thebiogrid.org/index.php | A curated database of genetic and protein-protein interactions for a number of organisms, including <i>S. cerevisiae</i> . |
| WoLF PSORT Protein Subcellular Localization Prediction | http://wolfsort.org/ | The latest version of pSORT, an on-line tool for the prediction of the sub-cellular localisation of proteins, including in yeast and other fungi. |
| Yeast transport proteins database (YTPdb) | http://homes.esat.kuleuven.be/~sbroehe/ytpdb | A manually annotated set of almost 300 <i>S. cerevisiae</i> proteins either demonstrated or predicted to be transporters. |
| KEGG (Kyoto Encyclopedia of genes and Genomes) | http://www.genome.jp/kegg/ | An invaluable source for metabolic pathway maps for <i>Saccharomyces</i> , <i>Pichia</i> , <i>Yarrowia</i> , and other yeast of biotechnological interest. |

| | | |
|--|---|---|
| Yeastnet (A consensus reconstruction of yeast metabolism) | http://www.comp-sys-bio.org/yeastnet | A stoichiometric model of the <i>S. cerevisiae</i> metabolic network constructed by the yeast research community. |
| DBD (Database of Biological Databases) | http://www.biodbs.info/ | Many relevant and useful on-line resources were excluded from this Table for space reasons. If it does not contain what you need, this site may lead you to it. |