Current Biology, Vol. 14, R392–R394, May 25, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.cub.2004.05.015

Evolutionary Genomics: Yeasts Accelerate beyond BLAST

Dispatch

Ken Wolfe

Two new genome sequences confirm that a whole genome duplication occurred in an ancestor of *Saccharomyces cerevisiae*. This left a legacy of about 500 pairs of duplicated genes, many of which contribute to this yeast's ability to ferment glucose anaerobically; a few have been evolving so quickly they retain almost no sequence similarity to each other.

Dubliners are fond of complaining that they can wait for an hour at a bus stop, and then suddenly three buses arrive together. Something similar is happening in yeast genomics, with the recent publications of the genome sequences of *Ashbya gossypii* [1] and *Kluyveromyces waltii* [2], and the imminent release of four others [3]. Together with sequences published last year [4,5], they provide comprehensive genomic coverage of a group of about a dozen fungi that spans roughly the same evolutionary breadth as the chordates (Figure 1).

The bakers' yeast *Saccharomyces cerevisiae* has long been a model organism in many areas of biology. The large community of researchers interested in yeast provided the impetus, and indeed much of the labor force, that led to *S. cerevisiae* becoming the first eukaryote to have its genome sequenced, in 1996 [6]. More recently, the broader group of fungi that includes *S. cerevisiae* — the hemiascomycetes — has in turn become a model system for studying evolutionary genomics and processes such as speciation [7–9]. The hemiascomycetes have many attractive features for comparative genomics. Their genomes are compact (9–20 megabases) with few introns or repetitive elements. The combination of decades of genetics research and recent large-scale approaches means that we now know more about the functions of genes in *S. cerevisiae* than in any other organism. More than 70% of its genes have been assigned a name, which indicates that at least some primary investigation into the gene's function has been done. Lastly, many of the so-called 'non-conventional' yeasts whose genomes are now being sequenced are of interest in their own right — either as pathogens (*Candida glabrata*, *C. albicans*) or for their industrial applications (*A. gossypii*, *K. lactis*, *Yarrowia lipolytica*) — which has led to the development of gene manipulation tools for these species.

One of the most striking findings from yeast comparative genomics is that a whole genome duplication occurred right in the middle of the evolutionary tree of the hemiascomycetes (Figure 1). This makes the yeasts a model system for studying the evolution of polyploids, as well as everything else. The hypothesis that *S. cerevisiae* is a paleopolyploid — a species that went through a polyploid phase but whose chromosomes subsequently became distinct again — was initially controversial [7,10], but has now been proven by the new data. *K. waltii* and *A. gossypii* are two species that split off from the lineage leading to *S. cerevisiae* before it became polyploid (Figure 1), and the order of genes along their chromosomes reads like a merger between genes from pairs of *S. cerevisiae* chromosomes, which is exactly the pattern expected from polyploidization followed by extensive gene loss [10,11]. The *K. waltii* and *A. gossypii* data show that the duplication event encompassed the whole genome of an ancestor of *S. cerevisiae*, so it was a polyploidization, as opposed to an aneuploidy or any other kind of large-scale duplication.

There is extensive colinearity of gene order among these yeast species once the 1:2 relationship caused by genome duplication is taken into account. Consequently, comparing *S. cerevisiae* to pre-duplication species (Figure 1) allows all the surviving gene pairs that were formed by genome duplication to be identified, with little or no ambiguity [1,2]. I have suggested that duplicate genes formed by polyploidy should be called 'ohnologs', after Susumu Ohno [12], to distinguish them from other kinds of paralogs, because they are all the same age [13]. There are about 500 pairs of ohnologs in *S. cerevisiae*, which indicates that the current genome, with about 5500 genes, was made by duplicating an ancestral 5000-gene genome, followed by loss of one copy at 90% of the loci, to leave the original set of genes plus about 500 extras.

The 500 surviving gene duplicates seem to comprise two groups. One group consists of those where there has been almost no sequence divergence between the ohnologs — indeed, they may have undergone homogenization by gene conversion [2,14] — and they seem to have been retained because there is a direct evolutionary advantage to having extra dosage of the genes. This group includes almost all the cytosolic ribosomal protein genes, as well as other translation genes such as elongation factors (about 100 pairs in all).

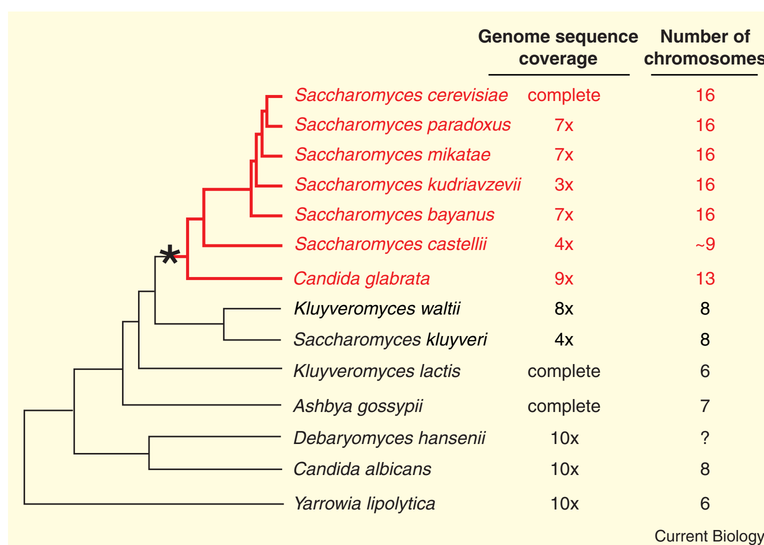
The second, larger, group is more interesting because it comprises pairs of genes that have diverged in sequence and often also in function. Kellis *et al.* [2] show that rates of sequence evolution in the latter group have often been asymmetric, with one ohnolog accumulating significantly more amino acid replacements than the other. They also argue that in many of these cases, the function of the slower-evolving copy is more similar to the ancestral function of the gene prior to duplication, and the faster-evolving copy has accelerated towards a derived function. This heterogeneity of rates may be the reason why phylogenetic analysis of ohnolog pairs produced a confusing picture of the timing of the genome duplication event [14].

Strikingly, some of the ohnolog pairs have very divergent sequences. The mean level of protein

Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland.
E-mail: khwolfe@tcd.ie

Figure 1. Approximate phylogenetic relationships among the sequenced yeast genomes.

The asterisk indicates the likely position of the whole genome duplication discussed in the text, and the species named in red are all descended from it. The 'pre-duplication' species, named in black, have roughly half the number of chromosomes of the post-duplication species.



sequence identity in ohnologs is 63% (similar to human versus pufferfish), but this average hides a very wide range of underlying evolutionary rates. Some examples of ohnologs with high divergence were already known, such as those encoding the Dig1 and Dig2 proteins, which inhibit the Ste12 transcription factor (18% protein sequence identity over their entire length in a ClustalW alignment).

Even more eye-opening are pairs of genes that completely fail to hit each other in BLASTP searches — their BLASTP Expect value is E > 1 — but which are inferred to be ohnologs because of their gene-order relationships. There are half-a-dozen such non-hitting pairs among the 500 pairs of ohnologs, and at least 21 pairs for which the mutual BLASTP hit is extremely weak ($E > 1x10^{-4}$). For example, Spo21 and Ysw1 have only 13% identity and do not hit each other in BLASTP, but their chromosomal gene contexts show unambiguously that they are ohnologs. They also have identical lengths (609 amino acids) and gene orientations, and they are both induced during meiosis. Similarly, the uncharacterised large proteins Yel025c and Yjr039w are ohnologs but have only 14% identity and do not hit each other in BLASTP.

In comparative genomics research, it is often assumed that homologous genes in groups of related species can be identified quite easily using BLAST searches, but these examples show that there is also an unseen world of homologs that are unrecognizable. Hence, many apparently species-specific 'orphan' genes may in fact be old loci whose sequences have become species-specific because they are changing so fast [4,7]. Of course, these genes are of great interest because they are likely to contribute to the phenotypic differences that set one species apart from another.

Looking at the lists of ohnologs compiled in the new studies, it is clear that the genome duplication played a direct role in the adaptation of the *S. cerevisiae* lineage towards fermentation [15]. *S. cerevisiae* differs from most other yeasts, including *K. waltii* and *A. gossypii*, by its ability to grow vigorously in the almost complete absence of oxygen, provided that glucose is available. Transcription of many ohnolog pairs is differentially regulated by oxygen, with one member of the pair being induced in hypoxic conditions whereas the other is aerobic [16]. Hypoxic growth is achieved by fermenting glucose to make ethanol and completely shutting down respiration — *S. cerevisiae* can even dispense with its mitochondrial genome under these conditions. What little oxygen is available is diverted into the essential purpose of making sterols and esters for membranes.

As well as contributing hypoxic isoforms of proteins, the genome duplication also greatly increased the ability of *S. cerevisiae* to respond to glucose [17]. The glucose-sensing pathway was split into high-affinity and low-affinity systems which detect different external concentrations of glucose and activate appropriate sets of hexose transporters, through the sensors Snf3 and Rgt2 and the repressors Mth1 and Std1, both of which are ohnolog pairs. The Snf1 pathway used by *S. cerevisiae* to repress gluconeogenesis and respiration in the presence of high concentrations of glucose was modified by the specialization of hexokinase into glucose specific (Hxk2) and non-glucose specific (Hxk1) ohnologs, and the formation of ohnologs Gal83 and Sip2 as alternative regulatory subunits of the Snf1 kinase. Likewise, the glucose-responsive protein kinase A pathway became considerably more complex when the Glk1/Emi2 glucokinases, Ras1/Ras2 G proteins, Tpk1/Tpk3 kinase subunits, and Msn2/Msn4 transcription factor pairs were formed by the genome duplication. It is remarkable to think that all these changes trace back to a single fortuitous event, and that natural selection must subsequently have acted simultaneously on all these loci to develop their modified functions.

Another area where genome duplication seems to have had a profound effect on *S. cerevisiae* biology is in the budding pattern, which in turn dictates colony morphology. Bud8 and Bud9, which are probably persistent tags marking the distal and proximal poles, respectively, of daughter cells [18], are ohnologs. This suggests that either the two poles of daughter cells are not differentially marked in pre-duplication species such as *K. waltii*,

or else a previous marking system was replaced by the Bud8/Bud9 system in *S. cerevisiae* following genome duplication. In fact, considering that most pre-duplication species grow as haploids, sporulating almost immediately after mating, whereas natural (homothallic) isolates of *S. cerevisiae* are diploid and form asci without conjugation [19], one could say that the *S. cerevisiae* lineage needed to invent a (bipolar) mitotic budding pattern for diploid cells. The switch from haploid to diploid as the major growth phase was made possible by the *HO* mating type switching gene, which appeared shortly before the genome duplication [20].

The publication of the *A. gossypii* [1] and *K. waltii* [2] genome sequences opens the door to detailed analysis of how *S. cerevisiae* evolved to become the glucose-fermenting machine it is today. Although complete genome sequences are invaluable, real experiments will also be necessary to give us a better understanding of the biology of pre-duplication species. Neither of the two newly sequenced species is ideal in this regard. *K. waltii* is almost completely uncharacterised and no laboratory tools have been developed for it. *A. gossypii* is an interesting species because it has no yeast phase — it can only grow in filaments — but genetic analysis is not possible so far because only one mating type has been discovered. It seems more likely that *K. lactis* or *S. kluyveri*, which are more familiar to many labs, will emerge as a model pre-duplication species. The latter is slightly more closely related to *S. cerevisiae* (Figure 1) but its genome has only been sequenced to four-times coverage so it has many gaps [5]. But it is important to bear in mind that all these pre-duplication species are themselves adapted to particular evolutionary niches and are in no sense more 'primitive' than *S. cerevisiae*, even if they do make lousy beer.

References

1. Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., Wing, R.A. Flavier, A., Gaffney, T.D., and Philippsen, P. (2004). The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science DOI: 10.1126/science.1095781.
2. Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature *428*, 617-624.
3. Génolevures-2 Sequencing Consortium feat. B. Dujon (2003). Comparative genomics of Hemiascomycetous yeasts: the systematic sequencing of *Candida glabrata*, *Kluyveromyces lactis*, *Debaryomyces hansenii*, and *Yarrowia lipolytica*. Yeast *20*, S15.
4. Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241-254.
5. Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., and Johnston, M. (2003). Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. Science *301*, 71-76.
6. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., *et al.* (1996). Life with 6000 genes. Science *274*, 563-567.
7. Souciet, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de Montigny, J., Dujon, B., *et al.* (2000). Genomic exploration of the Hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett *487*, 3-12.
8. Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., and Louis, E.J. (2000). Chromosomal evolution in *Saccharomyces*. Nature *405*, 451-454.
9. Delneri, D., Colson, I., Grammenoudi, S., Roberts, I.N., Louis, E.J., and Oliver, S.G. (2003). Engineering evolution to study speciation in yeasts. Nature *422*, 68-72.
10. Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. Nature *387*, 708-713.
11. Keogh, R.S., Seoighe, C., and Wolfe, K.H. (1998). Evolution of gene order and chromosome number in *Saccharomyces*, *Kluyveromyces* and related fungi. Yeast *14*, 443-457.
12. Ohno, S. (1970). Evolution by Gene Duplication. London: George Allen and Unwin.
13. Wolfe, K.H. (2001). Yesterday's polyploids and the mystery of diploidization. Nat. Rev. Genet. *2*, 333-341.
14. Langkjaer, R.B., Cliften, P.F., Johnston, M., and Piskur, J. (2003). Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. Nature *421*, 848-852.
15. Piskur, J. (2001). Origin of the duplicated regions in the yeast genomes. Trends Genet. *17*, 302-303.
16. Kwast, K.E., Lai, L.C., Menda, N., James, D.T., 3rd, Aref, S., and Burke, P.V. (2002). Genomic analyses of anaerobically induced genes in *Saccharomyces cerevisiae*: functional roles of Rox1 and other factors in mediating the anoxic response. J. Bacteriol. *184*, 250-265.
17. Geladé, R., Van de Velde, S., Van Dijck, P., Thevelein, J.M. (2003). Multi-level response of the yeast genome to glucose. Genome Biol. *4*, 233.
18. Harkins, H.A., Page, N., Schenkman, L.R., De Virgilio, C., Shaw, S., Bussey, H., and Pringle, J.R. (2001). Bud8p and Bud9p, proteins that may mark the sites for bipolar budding in yeast. Mol. Biol. Cell *12*, 2497-2518.
19. Kurtzman, C.P. (2003). Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the *Saccharomycetaceae*, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygotorulaspora*. FEMS Yeast Res. *4*, 233-245.
20. Butler, G., Kenny, C., Fagan, A. Kurischko, C., Gaillardin, C., and Wolfe, K.H. (2004). Evolution of the *MAT* locus and its Ho endonuclease in yeast species. Proc. Natl. Acad. Sci. USA *101*, 1632-1637.