



Parallel Evolution by Gene Duplication in the Genomes of Two Unicellular Fungi

Austin L. Hughes and Robert Friedman

Genome Res. 2003 13: 794-799

Access the most recent version at doi:[10.1101/gr.714603](https://doi.org/10.1101/gr.714603)

References This article cites 41 articles, 17 of which can be accessed free at:
<http://genome.cshlp.org/content/13/5/794.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center is a white button with the text "LEARN MORE". On the right is a woman wearing a red and white superhero cape and mask, with the Cellecta logo (a green molecular structure) and the word "CELLECTA" below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Parallel Evolution by Gene Duplication in the Genomes of Two Unicellular Fungi

Austin L. Hughes^{1,3} and Robert Friedman²

¹Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA; ²Department of Biology, Arizona State University, Tempe, Arizona 85287, USA

Phylogenetic analysis of conserved gene families in fission yeast *Schizosaccharomyces pombe* and brewer's yeast *Saccharomyces cerevisiae* showed that gene duplications have occurred independently in the same families in each of these two lineages to a far greater extent than expected by chance. These species represent distinct lineages of the phylum Ascomycota that independently evolved a "yeast" life cycle with a unicellular thallus that reproduces by budding, and many of the genes that have duplicated independently in the two lineages are known to be involved in crucial aspects of this life cycle. Parallel gene duplication thus appears to have played a role in the independent origin of similar adaptations in the two species. The results indicate that using phylogenetic analysis to test for parallel gene duplication in different species may help in identifying genes responsible for similar but independently evolved adaptations.

By giving rise to new genes encoding proteins of potentially novel function, gene duplication has long been thought to play a role in the adaptation of organisms to their environment (Ohno 1970; Hughes 1994; Lynch and Conery 2000). However, the patterns by which gene duplications have given rise to functionally new genes over the course of evolution remain poorly understood (Hughes 1999; Lynch and Conery 2000; Friedman and Hughes 2001a; Wolfe 2001; Gu et al. 2002; McLysaght et al. 2002). It has been proposed that duplication of different sets of genes in different lineages can promote speciation and adaptive radiation (Lynch and Force 2000). Thus, differential gene duplication is hypothesized to play a role in the origin of phenotypic diversity. On the other hand, it might be hypothesized that parallel duplication of the same set of genes can play a role in the adaptation of distantly related species to similar ecological niches. Thus, gene duplication may contribute to both divergent and convergent evolution at the phenotypic level.

We examined these processes by comparing the complete sets of protein-coding genes in two species of fungi in the phylum Ascomycota, the fission yeast *Schizosaccharomyces pombe* and brewer's yeast *Saccharomyces cerevisiae*. In contrast to many Ascomycota, both of these species have a life cycle characterized by a unicellular thallus that reproduces by budding (a "yeast" life cycle). *S. pombe* and *S. cerevisiae* belong to distinct classes (Archaeomycetes and Hemiacycomycetes, respectively) that have been separated for as long as 420 million years (Sipiczki 2000), and the yeast life cycle is a derived character in Ascomycota that has evolved independently in the two lineages (Berbee and Taylor 1993). Using phylogenetic analysis, we timed gene duplications in these two species relative to their last common ancestor to examine the extent to which gene duplication has occurred in a parallel manner in these two lineages that have adapted independently to a similar mode of life.

³Corresponding author.

E-MAIL austin@biol.sc.edu; FAX (604) 877-6085.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.714603>.

RESULTS

Phylogenetic Analyses

Assembling conserved protein families by homology search and constructing phylogenetic trees of each family by the quartet maximum likelihood method (Strimmer and von Haeseler 1996), we identified a total of 645 such families in which the phylogeny supported an orthologous relationship between a gene or genes of *S. pombe* and a gene or genes of *S. cerevisiae*. For each gene duplication in these families, branch order in the phylogenetic tree was used to test the hypothesis that gene duplication occurred prior to the last common ancestor of the two species (Fig. 1). A slightly higher proportion of duplications in *S. pombe* than in *S. cerevisiae* were dated prior to their last common ancestor, but the difference between the two species was not statistically significant (Fig. 1).

In 56 families, one or more gene duplications were found to have occurred independently in each species after their last common ancestor (Fig. 2). On the basis of the observed rates of gene duplication after the common ancestor in the 645 families examined, independent gene duplication in each species was observed far more frequently than expected by chance (Fig. 2). These 56 families included 22 families of ribosomal proteins. To test whether the observed pattern was specific to ribosomal proteins, we conducted the same analysis after excluding the ribosomal proteins from the data set. In this case also, independent gene duplications in the two species were observed in the same families to a far greater extent than expected by chance (Fig. 2).

Duplication Time Estimates

On the assumption of a constant rate of protein evolution ("molecular clock"), we dated 132 gene duplications occurring after the divergence of the two species relative to the age of their common ancestor, assumed to have occurred 420 million years ago (Sipiczki 2000). The genome of *S. cerevisiae* shows evidence of numerous duplicated genomic blocks, which are possibly the result of an ancient polyploidization event (Wolfe and Shields 1997; Seoighe and Wolfe 1999; Friedman and Hughes 2001a). In contrast, the genome of *S.*

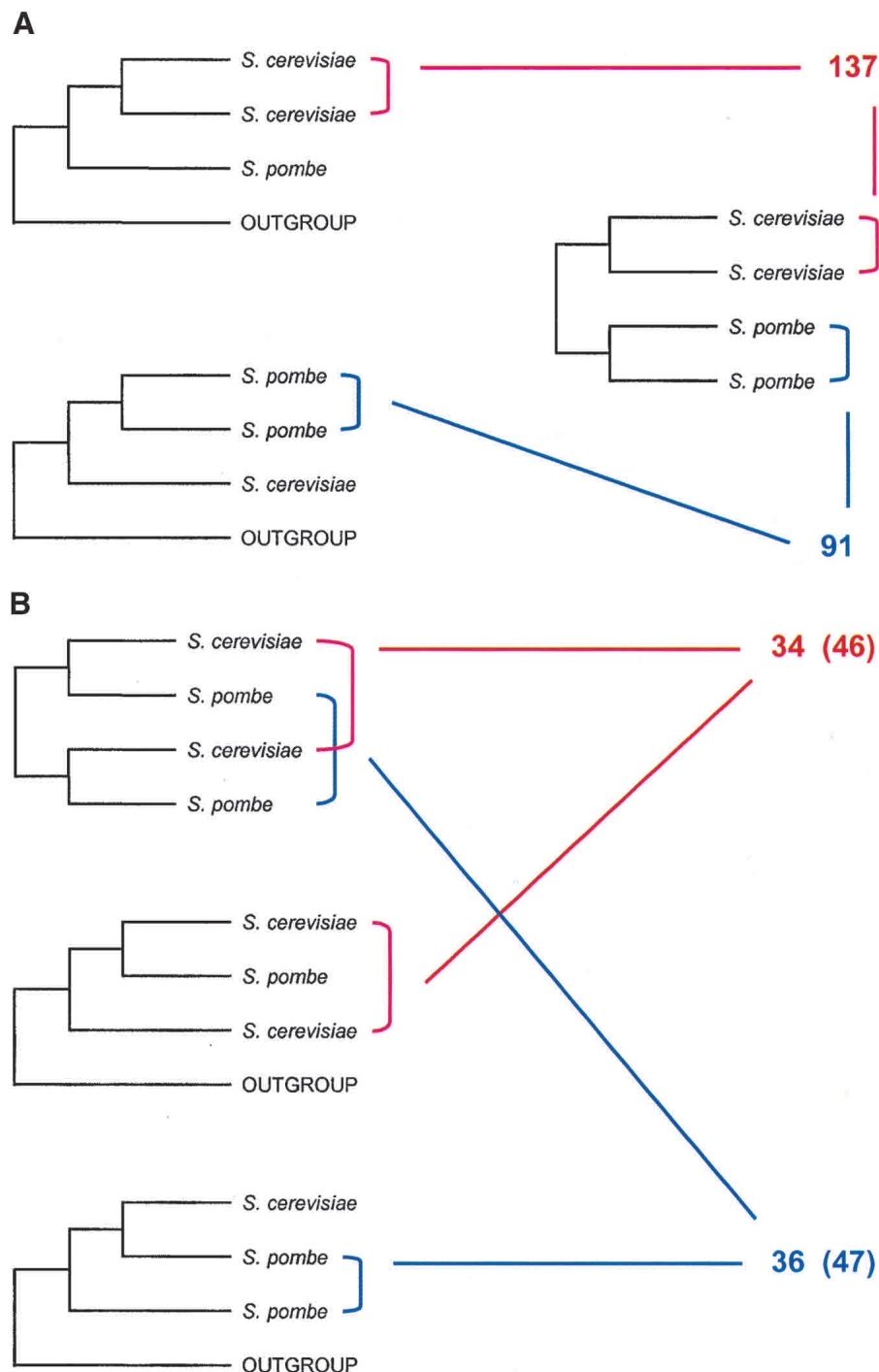


Figure 1 Summary of the results of phylogenetic analysis of protein families shared by *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, conducted by the quartet maximum likelihood method. (A) Numbers of duplications in each species that, according to the phylogeny, occurred after their last common ancestor. (B) Numbers of duplications that occurred before the last common ancestor of the two species, as supported by an internal branch (arrow) in the phylogenetic tree that was supported by $\geq 95\%$ of 10,000 quartet-puzzling steps. In parentheses are numbers of duplications for which the internal branch was supported by $\geq 70\%$ of 10,000 quartet-puzzling steps.

pombe shows no evidence of a duplicated structure (Wood et al. 2002). We dated gene duplications in *S. cerevisiae* involving elements of putative duplicated blocks (Seoighe and Wolfe 1999) separately from other duplications.

duplications in duplicated blocks of the *S. cerevisiae* genome were estimated to have occurred ~ 200 million years ago (Mya) is consistent with the occurrence of a polyploidization event around this time, as has been proposed previously (Friedman

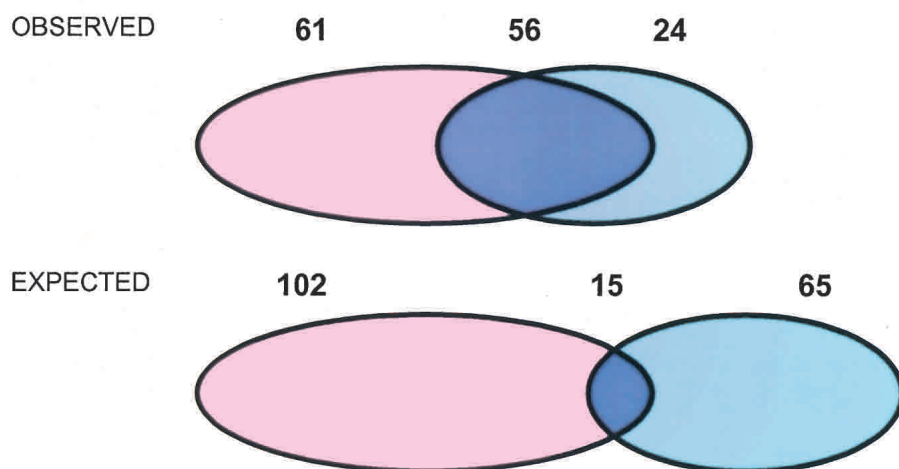
Elements of duplicated blocks showed a significantly lower variance in estimated duplication times than other pairs of duplicated genes in *S. cerevisiae* or *S. pombe* (Fig. 3). This result is consistent with the hypothesis that many of the former genes were duplicated within a narrow time frame as would occur in the event of polyploidization (Fig. 3). However, the range of duplication times even of genes within the blocks was so large as to cast doubt on the hypothesis that all of them were duplicated simultaneously. The estimated mean duplication times for *S. cerevisiae* genes in duplicated blocks (197 ± 12 million years) and not in duplicated blocks (205 ± 18 million years) were similar (Fig. 3). In contrast, estimated mean duplication time for *S. pombe* genes (310 ± 17 million years) was significantly earlier (Fig. 3).

DISCUSSION

The 56 gene families duplicated independently in *S. pombe* and *S. cerevisiae* include several proteins involved in cell growth and fission. In addition to ribosomal proteins (Venema and Tollervy 1999) already mentioned, these include numerous proteins known to be involved in bud site selection, cell wall biosynthesis, and other key aspects of the cell cycle (Table 1). Given the high frequency of genes with crucial cell cycle functions in the 56 independently duplicated families, it seems likely that gene duplication in these families was involved in adaptation of the two lineages to the unicellular yeast life cycle. The difference in mean duplication times between the two species indicates that adaptation to this life cycle occurred considerably earlier in the lineage of *S. pombe* than in that of *S. cerevisiae*. This is consistent with phylogenetic evidence supporting early branching of *S. pombe* within the Ascomycota and more recent radiation of the order Saccharomycetales (including *S. cerevisiae*; Berbee and Taylor 1993).

The fact that many of the gene duplications in duplicated blocks of the *S. cerevisiae* genome were estimated to have occurred ~ 200 million years ago (Mya) is consistent with the occurrence of a polyploidization event around this time, as has been proposed previously (Friedman

All proteins



Excluding ribosomal proteins

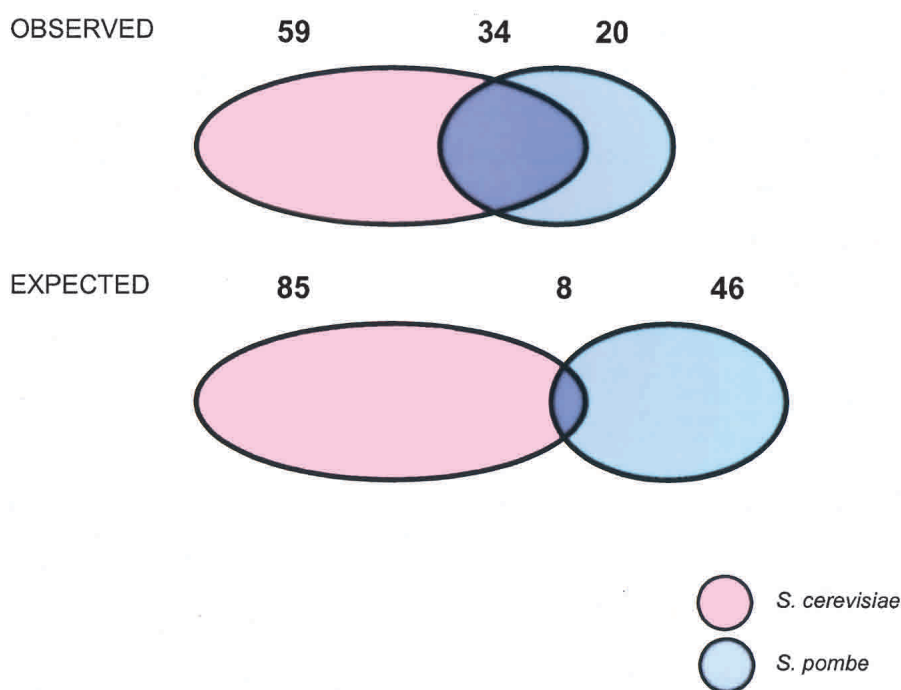


Figure 2 Observed numbers of gene families with one or more duplications after the last common ancestor of *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, illustrating the numbers of families duplicated in each species separately and in both species. Separate analyses were observed for a data set of all proteins (650 families) and for a data set excluding ribosomal proteins (623 families). Observed numbers were compared with expected numbers, calculated by multiplying the proportions of families with one or more duplications in *S. pombe* by the proportion of families with one or more duplications in *S. cerevisiae*. For both data sets the observed and expected numbers were significantly different ($\chi^2 = 117.2$ and 82.3 , respectively; $P < 0.00001$ in both cases).

and Hughes 2001a). Such a polyploidization event may have been a source of duplicate genes that this fungus was able to use in adapting to a unicellular lifestyle. It is of interest, however, that *S. pombe* has adapted to a similar lifestyle through

gene duplication without polyploidization. Furthermore, numerous gene duplications occurred in *S. cerevisiae* outside of the duplicated blocks and thus independently of the polyploidization event. Likewise, not all of the duplications within the blocks appear to have occurred at the same time (Fig. 3), indicating that duplication of genomic blocks has been an ongoing feature of the evolution of *S. cerevisiae* (Friedman and Hughes 2001a). Consequently, our results indicate that, although polyploidization can provide raw material for adaptive evolution, it is by no means essential to the origin of multiple new genes with novel functions. Thus, a peak of gene duplications over the course of evolutionary history need not imply a polyploidization event, contrary to the implication of some recent papers (e.g., Gu et al. 2002; McLysaght et al. 2002).

There is evidence that gene duplication occurs continually over the course of genome evolution (Lynch and Conery 2000). The mechanisms involved include “tandem” duplication of individual genes, duplication of larger genomic segments (Samonte and Eichler 2002), and (more rarely) whole genome duplication by polyploidization. Most duplicate genes are eventually lost, but duplicates that become specialized for adaptively important new functions have a much greater chance of being retained (Hughes 1994; Lynch et al. 2001). In each of the two yeast species, after their last common ancestor, substantial numbers of new duplicate genes have been retained in the genome. Our results indicate that such a peak of retained gene duplications is the signature of adaptive evolution at the genomic level.

Furthermore, our results indicate a novel approach for identifying genes potentially underlying phenotypic adaptations of interest. If two distantly related organisms have evolved similar phenotypic adaptations independently, phylogenetic analyses can be used as in the present study to identify gene

families that have duplicated independently in the two lineages. Such gene families will, in turn, provide a set of candidate genes for understanding the molecular basis of adaptations shared by the two lineages.

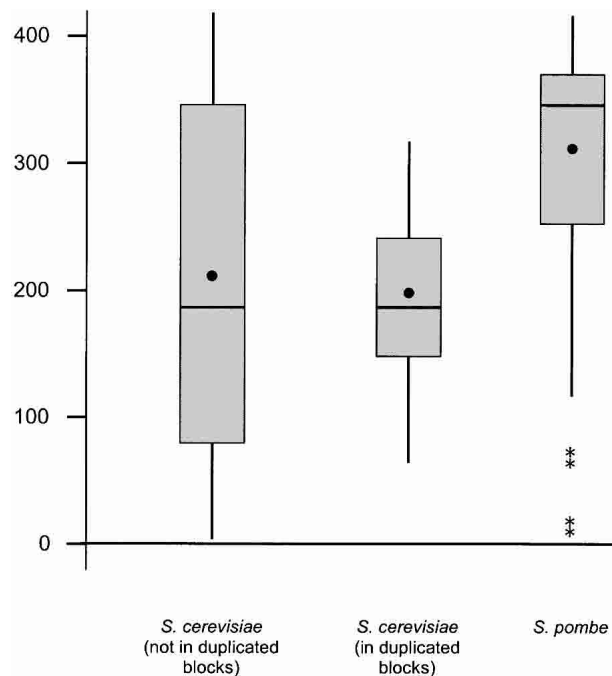


Figure 3 Box-and-whiskers plots of estimated duplication times for gene pairs duplicated after the last common ancestor of *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*, grouped as follows: (1) 48 duplications in *S. cerevisiae* of genes that are not elements of duplicated genomic blocks; (2) 39 duplications in *S. cerevisiae* of genes that are elements of duplicated genomic blocks; and (3) 46 duplications in *S. pombe*. The hypothesis of an equal variance in all three groups was rejected (Levene's test; $P = 0.004$). However, the hypothesis of equal variance in *S. cerevisiae* duplications not in blocks and *S. pombe* duplications could not be rejected (Levene's test). One-way analysis of variance showed a significant difference in mean duplication time between the three groups ($P < 0.001$), and the Kruskal-Wallis test showed a significant difference in median duplication times among the three groups ($P < 0.001$). In each plot, the box covers the range between the first (Q_1) and third (Q_3) quartiles in the data. A horizontal line indicates the median, and a dot the mean. The vertical lines ("whiskers") cover the range from $Q_1 - 1.5(Q_3 - Q_1)$ to $Q_3 + 1.5(Q_3 - Q_1)$. Outliers are indicated by asterisks.

METHODS

Sequences and Homology Searches

The set of complete protein translations (proteome) for *S. cerevisiae* was obtained at <http://genome-www.stanford.edu/Saccharomyces/>; that for *S. pombe* at http://www.sanger.ac.uk/Projects/S_pombe/; that for *Arabidopsis thaliana* at <http://www.tigr.org/>; that for *Caenorhabditis elegans* at http://www.sanger.ac.uk/C_elegans/; and that for *Drosophila melanogaster* at ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_sets. Proteome data sets from 38 complete genomes of bacteria and all available sequences from a set of other representative organisms (the slime mold *Dictyostelium discoideum*, the zebrafish *Danio rerio*, the pufferfish *Takifugu rubripes*, the frog *Xenopus laevis*, the human *Homo sapiens*, and the mouse *Mus musculus*) were obtained from the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>).

The text file of the nonredundant proteins was formatted as a database file using the BLAST tools obtained from the National Center for Biotechnology Information (NCBI) ftp site (<ftp://ftp.ncbi.nlm.nih.gov/>; Altschul et al. 1997). After the nonredundant proteome was determined for each genome,

each protein was used to search for homology among the rest of the proteome. This "all against all" BLAST method was performed using the blastall executable, which is packaged with the BLAST tools. Similarly, each protein in the nonredundant human proteome was searched against all proteins from other genomes. In searching human proteins against the remainder of the human proteome and in searching human proteins against nonhuman proteins, we used an expect value of $E = 10^{-50}$. A strict search criterion identifies as homologous only proteins showing evidence of homology throughout the length of the protein, rather than those with homology in only one domain or a few domains (Friedman and Hughes 2001a,b). In all cases, we used the defaults of a BLOSUM62 substitution matrix and the SEG filter (Wootton and Federhen 1993). The resultant records were filtered using MSPcrunch, a program to filter and convert the BLAST output to a tabular format (Sonnhammer and Durbin 1994).

Given all pairs of homologous proteins, a "single link" method was used to find the protein families. This step groups genes that share homology. For example, if gene A and B are in a family, and B and C are in another family, then A, B, and C are in a family. Furthermore, in this example, if A and D also share homology, then A, B, C, and D are in a family. Within families, amino acid sequences were aligned using CLUSTAL W 1.81 (Thompson et al. 1994).

Phylogenetic Analyses

Phylogenetic trees were constructed by the quartet maximum-likelihood (ML) method (Strimmer and van Haeseler 1996) as implemented in TREEPUZZLE 5.0, using the JTT (Jones et al. 1992) model of amino acid evolution and assuming that rate variation among sites followed a gamma distribution. On the basis of tree topology, we determined the time of each gene duplication event in *S. cerevisiae* or *S. pombe* relative to divergence of *S. cerevisiae* and *S. pombe*. This method of timing duplication events does not assume a constant rate of molecular evolution (molecular clock) and is independent of the rooting of the tree. We considered a branch to be significantly supported if it was supported in 95% or more of 10,000 puzzling steps; this represents a highly conservative test for significance of an internal branch (Strimmer and van Haeseler 1996). Strimmer and van Haeseler (1996) recommend the use of a 70% criterion; we therefore report these results as well (Fig. 1). We concluded that a gene duplicated prior to the cladogenetic event if the internal branch supporting that duplication was significantly supported.

Molecular Clock Analyses

To have an estimate of the time of duplication of genes in *S. cerevisiae* and *S. pombe* that duplicated after the last common ancestor of the two species, we applied the linearized tree method to these fungal sequences (Takezaki et al. 1995). The trees were calibrated using 420 million years as the divergence time of the two species (Sipiczki 2000). Note that the relative values of the estimated duplication times for genes duplicated within the *S. pombe* and *S. cerevisiae* lineages (Fig. 3) are not dependent on the accuracy of the estimate of the age of their common ancestor.

The Poisson model of amino acid evolution was used in constructing the linearized trees. This simple model was used because it has a lower variance than more complex models (Nei and Kumar 2000) and because, in these analyses involving relatively closely related sequences, it yielded essentially identical results to more complex models (data not shown). Ribosomal proteins were not used in the linearized tree analyses because amino acid sequences of these proteins were identical or nearly so within species, making a divergence time estimate impossible. Also excluded were other families in which the amino acid sequences within one species were identical. Furthermore, we excluded any case in which the

Table 1. Gene Families Independently Duplicated in *S. cerevisiae* and *S. pombe* Encoding Proteins With Known Roles in Processes Important to the Cell Cycle (Excluding Ribosomal Proteins)

<i>S. cerevisiae</i> duplicates	Cellular processes	References
<i>TOR1-TOR2</i>	Cell cycle progression	Powers and Walter 1999
<i>TRK1-TRK2</i>	Cell cycle progression, cell wall maintenance	Yenush et al. 2002
<i>PKH1-PKH2-YDR466W</i>	Actin cytoskeleton organization, cell wall maintenance	Schmelze et al. 2002
<i>ITR1-ITR2</i>	Lipid metabolism, mating and sporulation	Niederberger et al. 1998
<i>PRK1-ARK1-AKL1</i>	Actin cytoskeleton organization	Watson et al. 2001
<i>ROM1-ROM2</i>	Bud formation	Ozaki et al. 1996; Ni and Snyder 2001
<i>SKN1-KRE6</i>	Cell wall biosynthesis, bud site selection	Montijn et al. 1999; Ni and Snyder 2001
<i>SPO1-PLB1-PLB2-PLB3</i>	Lipid metabolism, meiosis	Fyrst et al. 1999
<i>HES1-KES1</i>	Posttranslational modification (Golgi)	Li et al. 2002
<i>SUR1-YBR161W</i>	Cell cycle progression	Helliwell et al. 1998
<i>FEN1-ELO1</i>	Lipid metabolism, bud site selection	Baudry et al. 2001; Ni and Snyder 2001
<i>ANB1-HYP2</i>	Protein synthesis	Mehta et al. 1990
<i>LAC1-LAG1</i>	Posttranslational modification (Golgi)	Schorling et al. 2001
<i>GGA1-GGA2</i>	Posttranslational modification (Golgi)	Black and Pelham 2000
<i>KTR5-KTR6</i>	Posttranslational modification (Golgi)	Lussier et al. 1999
<i>SIM1-SUN4</i>	Cell septation	Mouassite et al. 2000
<i>PTC2-PTC3</i>	Cell division (DNA checkpoint)	Marsolier et al. 2000
<i>RCK1-RCK2</i>	Inhibition of meiosis	Ramne et al. 2000

assumption of linearity was not met in the tree involving sequences from *S. pombe* and *S. cerevisiae*.

In separate preliminary analyses for each of the two species, we used *t*-tests to test the hypothesis of the equality of mean divergence time in the following two groups of genes: (1) genes duplicated in each species; (2) genes duplicated in only one of the species. In each species, there was no significant difference between the means for these two groups. For *S. pombe*, the mean estimated divergence time for genes duplicated independently in both species was 316 ± 18 (standard error) Mya, whereas that for genes duplicated only in *S. pombe* was 300 ± 33 Mya. For *S. cerevisiae*, the mean estimated divergence time for genes duplicated independently in both species was 192 ± 18 Mya, whereas that for genes duplicated in *S. cerevisiae* it was 206 ± 15 Mya. Because there was no significant difference between these two groups in each species, the two groups were pooled within species for subsequent analyses.

Estimated divergence times were compared by one-way analysis of variance among the following three groups: (1) *S. pombe* genes; (2) *S. cerevisiae* genes that are elements of putative duplicated blocks (as defined by Seoighe and Wolfe 1999); and (3) *S. cerevisiae* genes that are not elements of putative duplicated blocks. Levene's test of homogeneity of variance rejected the hypothesis of homogeneity of variance between these groups; however, the hypothesis of homogeneity of variance between groups 1 and 3 was not rejected. One-way analysis of variance was used in analyzing the data because it is known to be resistant to lack of homogeneity of variance (Lindman 1974).

ACKNOWLEDGMENTS

This research was supported by NIH grants R33 GM066710 and R01 GM43940 to A.L.H. We are grateful for comments by Federica Verra and Aoife McLysaght.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and

- PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Baudry, K., Swain, E., Rahier, A., Germann, M., Batta, A., Rondet, S., Mandala, S., Henry, K., Tint, G.S., Edlind, T., et al. 2001. The effect of erg26-1 mutation on the regulation of lipid metabolism in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**: 12702–12711.
- Berbee, M.L. and Taylor, J.W. 1993. Dating the evolutionary radiations of true fungi. *Can. J. Bot.* **71**: 1114–1127.
- Black, M.W. and Pelham. H.R.B. 2000. A selective transport route from Golgi to late endosomes that requires the yeast GGA proteins. *J. Cell Biol.* **151**: 587–600.
- Friedman, R. and Hughes, A.L. 2001a. Pattern and timing of gene duplication in animal genomes. *Genome Res.* **11**: 1842–1847.
- . 2001b. Gene duplication and the structure of eukaryotic genomes. *Genome Res.* **11**: 373–381.
- Fyrst, H., Oskouiian, B., Kuypers, F.A., and Saba, J.D. 1999. The PLB2 gene of *Saccharomyces cerevisiae* confers resistance to lysophosphatidylcholine and encodes a phospholipase B/lysophospholipase. *Biochemistry* **38**: 5864–5871.
- Gu, X., Wang, Y., and Gu, J. 2002. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat. Genet.* **31**: 205–209.
- Helliwell, S.B., Howald, I., Barbet, N., and Hall, M.N. 1998. TOR2 is part of two related signaling pathways coordinating cell growth in *Saccharomyces cerevisiae*. *Genetics* **148**: 99–112.
- Hughes, A.L. 1994. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B* **256**: 119–124.
- . 1999. *Adaptive evolution of genes and genomes*. Oxford University Press, New York, NY.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.
- Li, X., Rivas, M.P., Fang, M., Marchena, J., Mehrotra, B., Chaudhary, A., Feng, L., Prestwich, G.D., and Bankaitis, V.A. 2002. Analysis of oxysterol binding protein homologue Kse1p function in regulation of Sec14p-dependent protein transport from the yeast Golgi complex. *J. Cell Biol.* **157**: 63–77.
- Lindman, H.R. 1974. *Analysis of variance in complex experimental designs*. W.H. Freeman, San Francisco, CA.
- Lussier, M., Sdicu, A.M., and Bussey, H. 1999. The KTR and MNN1 mannosyltransferase families of *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1426**: 323–334.
- Lynch, M. and Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.
- Lynch, M. and Force, A. 2000. The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* **156**: 590–605.
- Lynch, M., O'Hely, M., Walsh, B., and Force, A. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**: 1789–1884.
- Marsolier, M.-C., Roussel, P., Leroy, C., and Mann, C. 2000.

- Involvement of the PP2C-like phosphatase Ptc2p in the DNA checkpoint pathways of *Saccharomyces cerevisiae*. *Genetics* **154**: 1523–1532.
- McLysaght, A., Hokamp, K., and Wolfe, K.H. 2002. Extensive genomic duplication during early chordate evolution. *Nat. Genet.* **31**: 200–204.
- Mehta, K.D., Leung, D., Lefebvre, L., and Smith, M. 1990. The ANB1 locus of *Saccharomyces cerevisiae* encodes the protein synthesis initiation factor eIF-4D. *J. Biol. Chem.* **265**: 8802–8807.
- Montijn, R.C., Vink, E., Müller, W.H., Verkleij, A.J., Van den Ende, H., Henrissat, B., and Klis, F.M. 1999. Localization of synthesis of β 1,6-glucan in *Saccharomyces cerevisiae*. *J. Bacteriol.* **181**: 7414–7420.
- Mouassite, M., Camougrand, N., Schwob, E., Demaison, G., Laclau, M., and Guérin, M. 2000. The “SUN” family: Yeast SUN4/SCW3 is involved in cell separation. *Yeast* **16**: 905–919.
- Nei, M. and Kumar, S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York, NY.
- Ni, L. and Snyder, M. 2001. A genomic study of the bipolar bud site selection pattern in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **12**: 2147–2170.
- Niederberger, C., Gräub, R., Schweingruber, A.-M., Fankhauser, H., Rusu, M., Poitelea, M., Edenharter, L., and Schweingruber, M.E. 1998. Exogenous inositol and genes responsible for inositol transport are required for mating and sporulation in *Schizosaccharomyces pombe*. *Curr. Genet.* **33**: 255–261.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer Verlag, New York, NY.
- Ozaki, K., Tanaka, K., Imamura, H., Hhara, T., Kameyama, T., Nonaka, H., Hirano, H., Matsuura, Y., and Takai, Y. 1996. Rom1p and Rom2p are GDP/GTP exchange proteins (GEPs) for the Rho1p small GTP binding protein in *Saccharomyces cerevisiae*. *EMBO J.* **15**: 2196–2207.
- Powers, T. and Walter, P. 1999. Regulation of ribosome biogenesis by the Rapamycin-sensitive TOR-signaling pathway in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **10**: 987–1000.
- Ramne, A., Bilsland-Marchesan, E., Erickson, S., and Sunnerhagen, P. 2000. The protein kinases Rck1 and Rck2 inhibit meiosis in budding yeast. *Mol. Gen. Genet.* **263**: 253–261.
- Samonte, R.V. and Eichler, E.E. 2002. Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* **3**: 65–72.
- Schmelze, T., Helliwell, S.B., and Hall, M.N. 2002. Yeast protein kinases and the RHO1 exchange factor TUS1 are novel components of the cell integrity pathway in yeast. *Mol. Cell. Biol.* **22**: 1329–1339.
- Schorling, S., Vallée, B., Barz, W.P., Riezman, H., and Oesterhelt, D. 2001. Lag1p and Lac1p are essential for the acyl-coA-dependent ceramide synthase reaction in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **12**: 3417–3427.
- Seoighe, C. and Wolfe, K.H. 1999. Updated map of duplicated regions in the yeast genome. *Gene* **238**: 253–261.
- Sipiczki, M. 2000. Where does fission yeast sit on the tree of life? *Genome Biol.* **1**: reviews1011.1–1011.4.
- Sonnhammer, E.L.L. and Durbin, R. 1994. A workbench for large scale sequence homology analysis. *Comput. App. Biol. Sci.* **10**: 301–307.
- Strimmer, K. and von Haeseler, A. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**: 964–969.
- Takezaki, N., Rzhetsky, A., and Nei, M. 1995. Phylogenetic test of the molecular clock and linearized tree. *Mol. Biol. Evol.* **12**: 823–833.
- Thompson, J.D., Higgins, D.G., and Gibson, T. 1994. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Venema, J. and Tollervey, D. 1999. Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **33**: 261–311.
- Watson, H.A., Cope, M.J.T.V., Grien, A.C., Drubin, D.G., and Wendland, B. 2001. In vivo role for actin-regulating kinases in endocytosis and yeast epsin phosphorylation. *Mol. Biol. Cell* **12**: 3668–3679.
- Wolfe, K.H. 2001. Yesterday’s polyploids and the mystery of diploidization. *Nat. Rev. Genet.* **2**: 334–341.
- Wolfe, K.H. and Shields, D.C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708–713.
- Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., et al. 2002. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**: 871–880.
- Wootton, J.C. and Federhen, S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. *Comp. Chem.* **17**: 149–163.
- Yenush, L., Mulet, J.M., Arino, J., and Serrano, R. 2002. The Ppz protein phosphatases are key regulators of K⁺ and pH homeostasis: Implications for salt tolerance, cell wall integrity and cell cycle progression. *EMBO J.* **21**: 920–929.

WEB SITE REFERENCES

- [ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_sets; *Drosophila melanogaster* proteome.](ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_sets;Drosophila_melanogaster_proteome)
- <http://www.ncbi.nlm.nih.gov/>; National Center for Biotechnology Information (NCBI).
- [http://genome-www.stanford.edu/Saccharomyces; *Saccharomyces cerevisiae* proteome.](http://genome-www.stanford.edu/Saccharomyces;Saccharomyces_cerevisiae_proteome)
- [http://www.sanger.ac.uk/C_elegans; *Caenorhabditis elegans* proteome.](http://www.sanger.ac.uk/C_elegans;Caenorhabditis_elegans_proteome)
- [http://www.sanger.ac.uk/Projects/S_pombe; *Schizosaccharomyces pombe* proteome.](http://www.sanger.ac.uk/Projects/S_pombe;Schizosaccharomyces_pombe_proteome)
- [http://www.tigr.org; *Arabidopsis thaliana* proteome.](http://www.tigr.org;Arabidopsis_thaliana_proteome)

Received August 15, 2002; accepted in revised form March 4, 2003.