Sciencexpress Research Article

Comparative Functional Genomics of the Fission Yeasts

Nicholas Rhind,[1]¶ Zehua Chen,[2] Moran Yassour,[3,4,5]|| Dawn A. Thompson,[3]|| Brian J. Haas,[2]|| Naomi Habib,[5,6]|| Ilan Wapinski,[3,7]|| Sushmita Roy,[3,8]|| Michael F. Lin,[8] David I Heiman,[2] Sarah K. Young,[2] Kanji Furuya,[9] Yabin Guo,[10] Alison Pidoux,[11] Huei Mei Chen,[12] Barbara Robbertse,[13]* Jonathan M. Goldberg,[2] Keita Aoki,[9] Elizabeth H. Bayne,[11]† Aaron M. Berlin,[2] Christopher A. Desjardins,[2] Edward Dobbs,[11] Livio Dukaj,[1] Lin Fan,[2] Michael G. FitzGerald,[2] Courtney French,[6] Sharvari Gujja,[2] Klavs Hansen,[14]‡ Dan Keifenheim,[1] Joshua Z. Levin,[2] Rebecca A. Mosher,[15]§ Carolin A. Müller,[16] Jenna Pfiffner,[2] Margaret Priest,[2] Carsten Russ,[2] Agata Smialowska,[17,18] Peter Swoboda,[17] Sean M. Sykes,[2] Matthew Vaughn,[14] Sonya Vengrova,[19] Ryan Yoder,[13] Qiandong Zeng,[2] Robin Allshire,[11] David Baulcombe,[15] Bruce W. Birren,[20] William Brown,[16] Karl Ekwall,[17,18] Manolis Kellis,[8,3] Janet Leatherwood,[12] Henry Levin,[10] Hanah Margalit,[6] Rob Martienssen,[14] Conrad A. Nieduszynski,[16] Joseph W. Spatafora,[13] Nir Friedman,[5,21] Jacob Z. Dalgaard,[19] Peter Baumann,[22,23,24] Hironori Niki,[9] Aviv Regev,[3,4,24]¶ Chad Nusbaum[2]¶

[1]Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA. [2]Broad Institute of Massachusetts Institute of Technology and Harvard, 320 Charles Street, Cambridge, MA 02141, USA. [3]Broad Institute of Massachusetts Institute of Technology and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA. [4]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [5]School of Computer Science and Engineering, Hebrew University, Jerusalem 91904, Israel. [6]Department of Microbiology and Molecular Genetics, Faculty of Medicine, Hebrew University, Jerusalem 91120, Israel. [7]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Alpert 536, Boston, MA 02115, USA. [8]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street 32-D510, Cambridge, MA 02139, USA. [9]Microbial Genetics Laboratory, Genetic Strains Research Center, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. [10]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. [11]Wellcome Trust Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, The University of Edinburgh, 6.34 Swann Building, Mayfield Road, Edinburgh EH9 3JR, UK. [12]Department of Molecular Genetics and Microbiology, Life Science, Room 130, State University of New York, Stony Brook, NY 11794, USA. [13]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA. [14]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA. [15]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK. [16]Centre for Genetics and Genomics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK. [17]Center for Biosciences, Department of Biosciences and Nutrition, Karolinska Institute, 141 38 Huddinge, Sweden. [18]Department of Life Sciences, Södertörns Högskola, 141 89 Huddinge, Sweden. [19]Warwick Medical School, University of Warwick, Gibbet Hill Campus, Coventry CV4 7AL, UK. [20]Broad Institute of Massachusetts Institute of Technology and Harvard, 301 Binney Street, Cambridge, MA 02141, USA. [21]Alexander Silberman Institute of Life Sciences, Hebrew University, Jerusalem 91904, Israel. [22]Stowers Institute for Medical Research, Kansas City, MO 64110, USA. [23]Department of Molecular and Integrative Physiology, University of Kansas Medical School, Kansas City, KS 66160, USA. [24]Howard Hughes Medical Institute.

*Present address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Department of Health and Human Services, 45 Center Drive, Bethesda, MD, 20892, USA.

†Present address: Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, UK.

‡Present address: Evolva Biotech A/S, Bülowsvej 25, 1870 Frederiksberg C, Denmark.

§Present address: The University of Arizona, The School of Plant Sciences, 303 Forbes Building, 1140 East South Campus Drive, Tucson, AZ 87421, USA.

||These authors made equivalent contributions.

¶To whom correspondence should be addressed. E-mail: nick.rhind@umassmed.edu (N.R.), aregev@broad.mit.edu (A.R.), chad@broadinstitute.org (C.N.)

The fission yeast clade—comprising *Schizosaccharomyces pombe*, *S. octosporus*, *S. cryophilus*, and *S. japonicus*—occupies the basal branch of Ascomycete fungi and is an important model of eukaryote biology. A comparative annotation of these genomes identified a near extinction of transposons and the associated innovation of transposon-free centromeres. Expression analysis established that meiotic genes are subject to

antisense transcription during vegetative growth, which suggests a mechanism for their tight regulation. In addition, trans-acting regulators control new genes within the context of expanded functional modules for meiosis and stress response. Differences in gene content and regulation also explain why, unlike the *Saccharomycotina*, fission yeasts cannot use ethanol as a primary carbon source. These analyses elucidate the genome structure and gene regulation of fission yeast and provide tools for investigation across the *Schizosaccharomyces* clade.

The fission yeast genus *Schizosaccharomyces* forms a broad and ancient clade within the Ascomycete fungi (Fig. 1A) with a distinct life history from other yeasts (*1*). Fission yeast grow preferentially as haploids, divide by medial fission rather than asymmetric budding, and have evolved a single-celled lifestyle independently from the budding yeasts (Saccharomycotina). Fission yeasts share important biological processes with metazoans, including chromosome structure and metabolism (relatively large chromosomes, large repetitive centromeres, heterochromatic histone methylation, chromodomain heterochromatin proteins, siRNA-regulated heterochromatin and TRF-family telomere binding proteins), G_2/M cell cycle control, cytokinesis, the mitochondrial translation code, the RNAi pathway, the signalosome pathway and spliceosome components with metazoans. These features are absent or highly diverged in budding yeast. In general, core orthologous genes in fission yeast more closely resemble those of metazoans than do those of other Ascomycetes (*2*). Fission yeasts have also evolved innovations in carbon metabolism, including aerobic fermentation of glucose to ethanol (*3*). This convergent evolution with the budding yeast *Saccharomyces cerevisiae* offers insight into the evolution of complex phenotypes.

S. pombe is widely used as a model for basic cell-biological processes and to study genes implicated in human disease. To better understand its evolution and natural history, we have compared the genomes and transcriptomes of *S. pombe*, *S. japonicus*, *S. octosporus* and *S. cryophilus*, which constitute all known fission yeasts.

Genome sequence and phylogeny. We sequenced and assembled the genomes of *S. octosporus*, *S. cryophilus* and *S. japonicus* using clone-based and clone-free whole-genome shotgun (WGS) approaches (table S1). Each genome is ~11.5 Mb in size. *S. octosporus* and *S. cryophilus* are 38% GC; *S. japonicus* is 44%. By comparison, the *S. pombe* genome is 12.5 Mb in size and 36% GC. We assembled the *S. octosporus* and *S. japonicus* scaffolds into 3 full-length chromosomes of similar quality to the finished *S. pombe* genome (Fig. 1B, figs. S1 and S2, and tables S2 and S3) and identified telomeric sequence using WGS data (*4*). Telomere-repeats in *S. japonicus* (GTCTTA), *S. octosporus* (GGGTTACTT) and *S. cryophilus* (GGGTTACTT) matched a one and a half repeat-unit sequence at the putative telomerase-RNA locus, similar to the configuration in *S. pombe* (GGTTAC) (*5*). Using these motifs, we extended the *S. japonicus* and *S. octosporus* chromosomes into subtelomeric and telomeric sequence (*4*).

We constructed a phylogeny of the *Schizosaccharomycetes* within *Ascomycota* (Fig. 1A and fig. S3) from 440 single-copy core orthologs, placing the monophyletic *Schizosaccharomyces* species as a basal sister group to the clade including the filamentous fungi (*Pezizomycotina*) and budding yeast (*Saccharomycotina*). We found an average amino acid identity of 55% between all 1:1 orthologs between *S. pombe* and *S. japonicus*, similar to that between humans and the cephalochordate amphioxus (table S4). For the most closely related species, *S. cryophilus* and *S. octosporus*, 1:1 orthologs share 85% identity on average, similar to humans and dogs. The genetic diversity within *S. pombe* is low. Comparing *S. pombe* 972 to WGS of *S. pombe* NCYC132 and *S. pombe var kambucha*, two phenotypically distinct strains, revealed less than 1% nucleotide difference between the three strains (fig. S4 and table S5).

Eradication of transposons and reorganization of centromere structure. Transposons and other repetitive sequences are thought to be crucial for centromeric function through the maintenance of heterochromatin (*6*). These sequences evolve rapidly, but the evolutionary relationship between centromeres, transposons and heterochromatin is unclear, in part because fungal centromeres have not generally been included in genome assemblies. The *S. japonicus* genome harbors 10 families of gypsy-type retrotransposons (*4*) (fig. S5 and table S6). Sequence divergence of their reverse transcriptases suggests that these transposon families predate the last common ancestor of the *Ascomycetes*. However, a dramatic loss of transposons occurred after the divergence of *S. japonicus*; *S. pombe* harbors two related retrotransposon, Tf1 and Tf2; *S. cryophilus* has a single related retrotransposon, Tcry1; *S. octosporus* contains no transposons, but contains sequences related to reverse transcriptase and integrase that may represent extinct transposons (fig. S5 and table S6).

The disappearance of transposons in the post-*S. japonicus* fission yeast species correlates with the

appearance of the *cbp1* gene family, suggesting a transition in the control of centromere function. In *S. pombe*, Cbp1 proteins bind centromeric repeats and are required for transposon silencing and genome stability (*7, 8*). Although described as orthologs of CENP-B, a human centromere-binding protein, Cbp1 proteins apparently evolved independently within the *Schizosaccharomyces* lineage from a domesticated Pogo-like DNA transposase (*9*). The appearance of the *cbp1* gene family also correlates with the switch from RNAi-mediated transposon silencing in *S. japonicus* (see below) to a Cbp1-based mechanism in *S. pombe*, suggesting that this shift to Cbp1-based transposon control allowed the eradication of most transposons from the fission yeast genomes, possibly by promoting recombinational deletion between LTRs (*8*). Furthermore, the *cbp1* family is evolving rapidly (fig. S6), suggesting that Cbp1-based transposon silencing is a *Schizosaccharomyces*-specific innovation that arose after the divergence of *S. japonicus*.

The loss of transposons was accompanied by a significant reorganization of chromosome architecture that conserves centromere function, suggesting evolution of novel centromere structures that compensate for the loss of transposons. In *S. japonicus*, transposons cluster next to telomeres and centromeres, as in metazoans (Fig. 1, B and C). In the other *Schizosaccharomycetes*, the subtelomeres and pericentromeres are also repetitive, but lack transposons (Fig. 1C). However, like *S. japonicus*, the centromeric and subtelomeric repeats are confined to pericentromeric and subtelomeric regions, respectively, with one exception ─ a centromeric repeat involved in transcriptional silencing at the *S. pombe* mating-type locus (*10*). We confirmed that the centromeres are heterochromatic by histone H3 lysine-9 methylation mapping (fig. S7), and by showing that the *S. japonicus* centromeres are functional by meiotic mapping (table S2).

Although centromeric repeats evolve rapidly, differing even between related strains (*11*), individual repeat sequences tend to be similar within strains (Fig. 1C). No similarity was observed between the centromeric repeats of *S. pombe*, *S. octosporus* or *S. cryophilus*. However, both *S. pombe* and *S. octosporus* centromeres contain repeated elements, highly similar between chromosomes, that are arrayed in a larger inverted repeat structure around a unique core sequence (Fig. 1C), suggesting that they are homogenized by non-reciprocal recombination. This contrasts with a lack of symmetry in *S. japonicus*, and implies that transposition occurs more rapidly than homogenization by recombination. Thus, the suppression of transposition likely led both to the degeneration of transposon sequences and to the evolution of symmetric centromeric repeats.

Despite the divergence of centromere sequence and of gene order on the chromosome arms, karyotype and pericentromeric gene order are conserved between *S. pombe* and *S. octosporus* (fig. S8). Thus, although gene conversion maintains the similarity of centromeric repeats between the different centromeres, crossover recombination between centromeres is suppressed. We observed neither centromeric translocations nor neo-centromere events within these lineages, despite the fact that centromeres can occur at novel locations in manipulated *S. pombe* strains. The retention of repetitive elements in the centromeres of (*12*) *S. pombe*, *S. octosporus* and *S. cryophilus*, even as they have lost their transposons, implies that centromeric repeats have an important function.

Since siRNAs are involved in both transposon silencing and centromere function (*13*), we investigated these roles in the *Schizosaccharomyces* lineage. In *S. pombe*, the centromeric repeats produce dicer-dependent siRNAs required for maintenance of centromeric structure, function and transcriptional silencing via Argonaute-dependent heterochromatin formation (*14*). However, transposons are silenced in *S. pombe* by RNAi-independent mechanisms and do not produce abundant siRNAs (Fig. 1B and fig. S9) (*7*). To investigate whether centromere-directed siRNA production is conserved within the transposon-rich centromeres of *S. japonicus*, we sequenced small RNAs from log-phase *S. japonicus* cultures (which have a modal size of 23 nt) (*4*) and found that 94% map to transposons, both telomeric and centromeric (Fig. 1B and fig. S9). The fact that siRNAs map to transposons in *S. japonicus* but not in *S. pombe* suggests that either the fission yeast RNAi pathway targets repetitive sequences instead of mobile elements *per se*, or that the pathway evolved away from an ancestral role in transposon control to a dedicated role in heterochromatin function.

Evolution of mating-type loci. The structure of the mating-type loci and the *cis*-acting elements that regulate mating-type switching is highly conserved across all four species (fig. S10). The expressed *mat1* locus can contain either the plus (P) or the minus (M) allele and switches between the two by epigenetically-programmed gene conversion (*15–17*) from two heterochromatically-silenced donor cassettes: *mat2-P* and *mat3-M* (figs. S10 and S11). *cis*-acting regulatory sequences required for epigenetic imprinting and recombinational switching (*18–20*) are conserved (*4*) (fig. S11), as is the epigenetically-programmed genomic mark associated with *mat1* (*15*).

In contrast, none of the *cis*-acting sequences involved in transcriptional repression of the silent cassettes in *S. pombe* are identifiable in the other species, although the donor cassettes are enriched for H3K9me heterochromatin (fig. S7). In *S. japonicus*, the silent *mat* loci directly abut the centromere of Chromosome 3, suggesting that they may be silenced by a positional effect. In *S. octosporus* and *S. cryophilus*, the *mat* loci are distant from the centromeres, but each contains a conserved region of transposon remnants, which may be silencing triggers. Interestingly, they also contain inverted repeats, albeit shorter and less similar to each other than the inverted repeats that flank the *mat2/3* locus in *S. pombe* (*21*). Thus, their silencing strategies may share elements from both *S. pombe* and *S. japonicus*. These results suggest that the mechanisms of imprinting and switching have been conserved, but that the strategies for establishing heterochromatin are plastic.

Comparative annotation of transcriptomes. We annotated the three genomes using standard methods and compared them with *S. pombe* (*4*). We then deep-sequenced polyA-enriched, strand-specific cDNA (*22–24*) (RNA-Seq), and constructed *de novo* transcript models (fig. S12) for log phase, glucose depletion, early stationary phase and heat shock from *S. pombe*, *S. octosporus*, and *S. japonicus* and log phase, glucose depletion and heat shock from *S. cryophilus*.

In *S. pombe*, we reconstructed 4277 out of 5064 previously annotated genes; of the remaining 788 genes, 60% were covered over at least 90% of their length. 400 of our transcript models change coding exon structure of the gene, 95% of which maintained or improved conserved coding capacity (tables S7 and S8 and fig. S12) (*25*). In addition, we identified 253 UTR introns. Lastly, we found 89 new protein-coding genes in *S. pombe*, 53 of which are conserved (table S7 and fig. S13). We found no evidence that intron-rich fission yeast genes engage in metazoan-like alternative splicing (*26*). We found evidence for 433 alternative splicing events in *S. pombe* in the form of intron retention and alternative splice-donor or -acceptor usage, but no evidence of exon skipping or alternative exons; we found similar levels of splice variants in the other species (table S9). However, since many of these variants disrupt the coding capacity (figs. S14 and S15) and only a minority of intron-retentions (146/393) are conserved between two or more species, we suspect that much of alternative splicing in fission yeast represents nonproductive splicing variants. Interestingly, in some cases the non-spliced variant may be the protein-coding isoform (figs. S14, C and D, and S15, and table S10).

Transcription primarily represents protein-coding transcripts. The majority of stable fission yeast transcripts originate from annotated protein coding genes. Most of the *S. pombe* genome is transcribed (*22*) with 91% of nucleotides covered by at least one RNA-Seq read. However, most transcription, as measured by steady-state polyA-enriched RNA levels, is associated with well-defined transcripts, most of which are protein coding. Specifically, 37% of intergenic nucleotides (between the UTRs of annotated protein coding transcripts) are not detectibly expressed and 90% of transcribed intergenic nucleotides account for only 0.16% of the polyA-enriched transcript signal. Moreover, the median expression level of exonic sequence (99.1% of which are detectably expressed) is 305-fold higher than that of intergenic sequence (Fig. 2 and table S11), with intergenic transcription enriched within origins of DNA replication (fig. S16) – gene-free loci with nucleosome-free regions (*27–29*) that may provide permissive loci for ectopic transcriptional initiation (*30*).

Transcription of coding genes is heavily biased to the sense strand. Of coding genes, 73% have less than 5% of their RNA-Seq reads on the antisense strand. Genes with more than 5% antisense reads are enriched for convergent transcripts with intergenic distances of less than 200 bp ($p < 10^{-8}$, hypergeometric test), but not with those of greater than 200 bp ($p > 0.1$), suggesting that much antisense transcription is due to read through of 3′-termination sites (*31*) (fig. S17). Thus, stable transcripts in fission yeast genomes are primarily associated with known transcription units. We discuss notable exceptions below.

Conservation of gene content and structure. Despite the evolutionary breadth of the fission-yeast clade, as measured by amino-acid divergence, their gene content and structure are remarkably conserved. Of ~5000 coding genes in fission yeast species, 4218 are 1:1:1:1 orthologs across the clade, with the remainder of the orthologous groups containing genes that have been duplicated or deleted since their last common ancestor (Table 1 and fig. S12). Protein kinases are even more conserved in gene content; 93% (102/110) of *S. pombe* protein kinases are 1:1:1:1 orthologs (*4*). Moreover, of 3601 *S. pombe* introns in 2616 spliced 1:1:1:1 orthologs, 2901 (81%) are identical across the four species (table S13). The majority of changes are due to the gain of species- and clade-specific genes (tables S12 and S14) (*4*). Overall, the conservation of gene content, gene order and gene structure within *Schizosaccharomyces* is higher than expected given the level of amino acid divergence. From amino acid divergence, we estimate that the fission yeast clade arose about 250 million years ago (fig. S3).

However, the conservation of gene content is significantly higher than that within *Saccharomyces* or *Kluveromyces*, both of which have much lower amino acid divergence (table S15), suggesting that fission yeast amino acid sequences are evolving anomalously quickly, or that genome structures are unusually stable.

The majority of gene changes are due to the gain of species- and clade-specific genes (table S12). We tested whether gene gain is due to rapid divergence of orthologous genes by looking for co-linearity in regions with species-specific genes, and examined these regions for signs of sequence similarity. We found that 94/317 *S. pombe*-specific genes are in the same position relative to neighboring genes as genes specific to other species (table S16). Of these, 9 show greater than 15% identity to a cognate gene in another species, suggesting that they are rapidly diverged orthologs (*4*).

We also found 34 *S. pombe* candidates for horizontal gene transfer from bacteria, including two published examples (*4, 32, 33*) (table S17), and similar numbers in the other species. Of these, 16 appear to have occurred before the radiation of the clade, and 9 appear to be specific to *S. pombe*.

Evidence for intergenic and antisense non-coding transcripts. We identified 1097 putative transcript models in *S. pombe* supported by strand-specific RNA-Seq data but containing no obvious coding capacity and having no correspondence to well-defined non-coding RNAs (*22, 24, 34*) (fig. S18, and tables S18 and S19). Of these potential ncRNAs, 449 are intergenic and 648 are antisense, overlapping a coding gene on the other strand by at least 30%. 213 of the ncRNAs overlap an annotated UTR on the same strand, suggesting that they may be alternative UTRs. Nevertheless, the data support 338 of the intergenic and 546 of the antisense ncRNAs as distinct transcripts (*4*).

Of the 338 distinct intergenic ncRNAs in *S. pombe*, 138 are conserved in location in at least one other species (table S41). Moreover, 26 of the intergenic ncRNAs are conserved in sequence and of these, 9 are conserved in both location and sequence, suggesting they represent potentially biologically important noncoding RNAs. The transcripts that are conserved in location but not in sequence may represent functional transcripts that have diverged beyond recognition. Of the antisense transcripts, 328 (51%) are conserved across two or more genomes (table S20), suggesting that they are biologically significant (*35*).

Antisense regulation of meiotic transcription. Across fission yeast, the ~250 genes with greater antisense transcription than sense transcription (table S21) are significantly enriched for meiotic genes ($p = 10^{-10}$ for *S. pombe*, hypergeometric test) (Fig. 3, fig. S19, and tables S22 and S23), consistent with observations in *S. pombe* and *S. cerevisiae* (*24, 35*). Several antisense-transcribed genes have been proposed to be regulated by intron retention (*36, 37*), however these studies did not use strand-specific approaches, making it impossible to distinguish unspliced sense transcripts from antisense transcripts. We find no evidence of alternative splicing of any of these genes.

Antisense transcription of meiotic genes does not uniformly decrease as cognate sense transcription increases during meiosis (fig. S20). This observation suggests that antisense transcription does not inhibit sense transcription, in contrast to the anti-correlation observed in *S. cerevisiae* (*30, 35*). Furthermore, meiotic genes are not enriched among genes with greater than 5% antisense transcription but less than 100% antisense transcription ($p = 0.47$, hypergeometric test), consistent with a stoichiometric mechanism of regulation in which antisense transcripts directly bind to and inhibit the stability or translation of sense transcripts.

Global conservation of expression programs within fission yeasts. To identify conserved modules of co-expressed genes, we examined expression patterns across the four conditions and between the four fission yeast with phylogenic clustering (Fig. 4). We found that patterns of gene expression between species grown in similar conditions are generally conserved, with dominant patterns associated with growth (log and heat shock) and stress (glucose depletion and early stationary phase). Moreover, similar expression clusters are enriched for similar gene annotations across the species.

Fission yeast up regulate genes involved in mitosis, including those involved in the kinetocore, the spindle pole body and the anaphase-promoting complex, in response to glucose depletion (table S24). In contrast, several classes of genes involved in growth are down regulated (*4*). None of these genes is significantly regulated in glucose depletion in *S. cerevisiae* (*38*).

***cis*-regulatory mechanisms are associated with novel and expanded functions.** Promoter motifs with conserved regulatory function across *Ascomycota* show new functionality among the *Schizosaccharomyces*. For example, the motif bound by Rtg3 in *S. cerevisiae* is associated with amino acid metabolism genes across the phylum. In fission yeast however, it is also enriched in genes responsive to various stress responses (Fig. 5A). Of the stress genes that have Rtg3 motifs in *S. pombe*, 36% are found only in the *Schizosaccharomyces* clade, and many are also associated with the Atf1 motif, a conserved

regulator of the stress response (Fig. 5B). Rtg3 does not have a detectable ortholog in the *Schizosaccharomyces* clade (*39*), but the motif recognized by Rtg3 in *S. cerevisiae* is clearly identifiable in fission yeast, suggesting that these regulatory motifs are more conserved than their binding proteins. We also found a similar acquisition of *Schizosaccharomyces*-specific genes by the Fkh1- and MBF-associated motifs, which regulate meiotic transcription in *S. pombe* (*4, 40, 41*). In particular, these two motifs were enriched in genes with antisense transcripts (Fig. 5C). Most of the Fkh1/Mei4 target genes with antisense transcripts (80%, 47 genes) are meiotic genes, the majority of which are specific to the *Schizosaccharomyces* clade (Fig. 5C).

Gene content reflects glucose-dependent lifestyle. Fission yeast and budding yeast of the *Saccharomyces* clade independently evolved the ability to produce ethanol by aerobic fermentation (*3, 42*). In contrast to the convergent evolution of ethanol production, the utilization of ethanol has not converged; although budding yeast can efficiently catabolize ethanol, fission yeast cannot use ethanol as a primary carbon source. The evolution of aerobic fermentation in budding yeast involved changes in gene content, most notably following a whole genome duplication (WGD) event, and in regulatory mechanisms of glucose repression (*3, 43*).

Like budding yeast, fission yeast have duplicate copies of the pyruvate decarboxylase (*pdc*) gene, needed to funnel pyruvate to fermentation. They also have orthologs of several activators and repressors of respiratory genes, including Hap2/3/4/5 complex members, the Adr1, Tup and Mig transcriptional regulators, and the Snf1-Sip1/2 kinase (*3*). However, there are substantial distinctions in gene content between fission yeast and the post-WGD budding yeast (fig. S21). We identified loss of the glyoxylate cycle, loss of the glycogen biosynthesis, fewer glycolytic paralogs, loss of the gluconeogenic enzyme phosphoenolpyruvate carboxykinase, lack of expanded *adh* genes, and lack of transcriptional regulators of glucose repression as differences that illuminate the distinct metabolic capacities of fission yeast (*4*). All of these adaptations are consistent with the inability of fission yeast to consume ethanol as a sole carbon source. The loss of conserved enzymes highlights how fission yeast came to depend solely on glucose.

In both fission yeast and budding yeast, as glucose is depleted the expression of respiratory genes (oxidative phosphorylation enzymes, TCA cycle) is induced. However, unlike *S. cerevisiae* (*38*), in fission yeast the expression of the genes encoding the pyruvate dehydrogenase complex and *adh1* are reduced, preventing the efficient use of pyruvate for respiration. Instead, the expression of the *ald* genes is induced, which may provide an alternative mechanism for generating acetyl-coA in fission yeast.

Thus, the lack of efficient ethanol catabolism by fission yeast demonstrates that aerobic fermentation did not evolve to create a consumable by-product. Instead, ethanol is a waste product, possibly produced because it is toxic to competing micro-organisms. Interestingly, aerobic fermentation appears to have evolved as early as 200 million years ago in fission yeast (fig. S3), long before the WGD and subsequent evolution of aerobic fermentation in budding yeast.

Conclusions. Our comparative analysis of genome structure and expression in the fission yeast, especially the analysis of centromere structure and evolution, demonstrates how chromosomal features can be rearranged while retaining function and maintaining stable positions across taxa. We also provide insight into centromeric biology and elucidate conserved antisense transcription that may play a systematic role in meiotic gene regulation. Lastly this study informs on the major evolutionary innovation of aerobic alcohol fermentation in microbial metabolism that arose in parallel in the fission yeast and budding yeast lineages. As these results demonstrate, comparative analyses improve the power of fission yeast as a model for eukaryotic biology.

References and Notes

1. S. L. Forsburg, *Trends Genet* **15**, 340 (1999).
2. V. Wood *et al.*, *Nature* **415**, 871 (2002).
3. C. L. Flores, C. Rodriguez, T. Petit, C. Gancedo, *FEMS Microbiol Rev* **24**, 507 (2000).
4. Materials and methods are available as supporting material on *Science* Online.
5. J. Leonardi, J. A. Box, J. T. Bunch, P. Baumann, *Nat Struct Mol Biol* **15**, 26 (2008).
6. L. H. Wong, K. H. Choo, *Trends Genet* **20**, 611 (2004).
7. H. P. Cam, K. Noma, H. Ebina, H. L. Levin, S. I. Grewal, *Nature* **451**, 431 (2008).
8. M. Zaratiegui *et al.*, *Nature* (2010).
9. C. Casola, D. Hucks, C. Feschotte, *Mol Biol Evol* **25**, 29 (2008).
10. S. I. Grewal, A. J. Klar, *Genetics* **146**, 1221 (1997).
11. B. Fishel, H. Amstutz, M. Baum, J. Carbon, L. Clarke, *Mol Cell Biol* **8**, 754 (1988).
12. K. Ishii *et al.*, *Science* **321**, 1088 (2008).
13. S. I. Grewal, *Curr Opin Genet Dev* **20**, 134 (2010).
14. T. A. Volpe *et al.*, *Science* **297**, 1833 (2002).
15. D. H. Beach, *Nature* **305**, 682 (1983).
16. R. Egel, *Curr Genet* **8**, 205 (1984).

17. S. Vengrova, J. Z. Dalgaard, *Genes Dev* **18**, 794 (2004).
18. S. Sayrac, S. Vengrova, E. L. Godfrey, J. Z. Dalgaard, *PLoS Genet* **7**, e1001328 (2011).
19. M. Kelly, J. Burke, M. Smith, A. Klar, D. Beach, *EMBO J* **7**, 1537 (1988).
20. B. Arcangioli, A. J. Klar, *EMBO J* **10**, 3025 (1991).
21. G. Singh, A. J. Klar, *Genetics* **162**, 591 (2002).
22. B. T. Wilhelm *et al.*, *Nature* **453**, 1239 (2008).
23. J. Z. Levin *et al.*, *Nat Methods* **7**, 709 (2010).
24. T. Ni *et al.*, *PLoS One* **5**, e15271 (2010).
25. M. F. Lin, I. Jungreis, M. Kellis, *Nature Precedings* (2010).
26. N. F. Kaufer, J. Potashkin, *Nucleic Acids Res* **28**, 3003 (2000).
27. M. Gomez, F. Antequera, *EMBO J* **18**, 5683 (1999).
28. M. L. Eaton, K. Galani, S. Kang, S. P. Bell, D. M. Macalpine, *Genes Dev* **24**, 748 (2010).
29. A. B. Lantermann *et al.*, *Nat Struct Mol Biol* **17**, 251 (2010).
30. Z. Xu *et al.*, *Nature* **457**, 1033 (2009).
31. M. Zofall *et al.*, *Nature* **461**, 419 (2009).
32. T. Matsuzawa *et al.*, *Appl Microbiol Biotechnol* **87**, 715 (2010).
33. T. Uo, T. Yoshimura, N. Tanaka, K. Takegawa, N. Esaki, *J Bacteriol* **183**, 2226 (2001).
34. N. Dutrow *et al.*, *Nat Genet* **40**, 977 (2008).
35. M. Yassour *et al.*, *Genome Biol* **11**, R87 (2010).
36. A. Moldon *et al.*, *Nature* **455**, 997 (2008).
37. N. Averbeck, S. Sunder, N. Sample, J. A. Wise, J. Leatherwood, *Mol Cell* **18**, 491 (2005).
38. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
39. I. Wapinski, A. Pfeffer, N. Friedman, A. Regev, *Nature* **449**, 54 (2007).
40. N. F. Lowndes, C. J. McInerny, A. L. Johnson, P. A. Fantes, L. H. Johnston, *Nature* **355**, 449 (1992).
41. H. Abe, C. Shimoda, *Genetics* **154**, 1497 (2000).
42. J. Piskur, E. Rozpedowska, S. Polakova, A. Merico, C. Compagno, *Trends Genet* **22**, 183 (2006).
43. M. Kellis, N. Patterson, M. Endrizzi, B. Birren, E. S. Lander, *Nature* **423**, 241 (2003).
44. J. Mata, R. Lyne, G. Burns, J. Bahler, *Nat Genet* **32**, 143 (2002).

Acknowledgments. Assemblies and annotations are available at GenBank (*S. octosporus*: ABHY04000000, *S. cryophilus*: ACQJ02000000, *S. japonicus*: AATM02000000), the Broad Institute *Schizosaccharomyces* website <http://www.broadinstitute.org/annotation/genome/schizosaccharomyces_group>, which provides search and visualization tools and pomBase <http://www.pombase.org>. The RNA-Seq and SNP data are at the NCBI SRA (see table S42). The *S. japonicus* siRNA datasets are at NCBI GEO as GSE26902 and GSE27837. This work was supported by NHGRI. C.N. and M.F.L. were supported by NHGRI; M.Y. was supported by a Clore Fellowship; I.W. is the HHMI fellow of the Damon Runyon Cancer Research Foundation; S.R. was supported by NSF; R.M. was supported by NIH; K.H. was supported by DRC; C.A.N and C.A.M were supported by BBSRC; P.B. was supported by the Stowers Institute and HHMI; Y.G. and H.L. were supported by NICHD; M.K was supported by the NIH, an NSF CAREER award, and the Sloan Foundation; A.R. was supported by HFSP, a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, the Sloan Foundation, an NIH Director's PIONEER award and HHMI. We thank the Broad Institute Sequencing Platform, A. Fujiyama and A. Toyoda for generating DNA sequence, M. Lara and N. Stange-Thomann for developing molecular biology protocols, J. Robinson, M. Garber, P. Muller for technical advice and support, A. Klar for providing *S. pombe var kambucha* (SPK1820), L. Gaffney for assistance with the figures, K. Mar and J. Mwangi for administrative support, and C. Cuomo for comments on the manuscript

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1203357/DC1
Materials and Methods
Figs. S1 to S25
Tables S1 to S42
References

26 January 2011; accepted 7 April 2011
Published online 21 April 2011; 10.1126/science.1203357

Fig. 1. *Schizosaccharomyces* phylogeny and chromosome structure. (**A**) A maximum-likelihood phylogeny of 12 fungal species from 440 core orthologs (each occurring once in each of the genomes) from fly to yeast. A maximum-parsimony analysis produces the same topology. Both approaches have 100% bootstrap support for all nodes. (**B**) The chromosome structure of *S. pombe, S. octosporus* and *S. japonicus*. The middle bar in each figure represents the chromosome and its centromere: red for Chromosome 1, blue for Chromosome 2 and yellow for Chromosome 3. Above and below each chromosome are depicted the chromosomes in the other two species to which the genes on the chromosome of interest map, using the same color scheme. Above the *S. pombe* and *S.*

japonicus chromosomes are depicted the distributions of transposons and mapping of siRNAs. *S. cryophilus* is not included because its genome has not been assembled into complete chromosomes. (**C**) The centromeric repeat structures of *S. pombe, S. octosporus* and *S. japonicus*.

Fig. 2. Polyadenylated transcription is predominantly confined to protein coding genes. The *S. pombe* genome was divided in to five different feature classes: protein coding sequence, intron sequence, untranslated sequence (5′ and 3′ UTRs) and intergenic sequence (all nucleotides between UTRs of protein coding genes). The frequency of RNA-Seq reads was calculated over sequential 20 bp windows across these features; for coding sequence, the frequency of antisense reads was also calculated. Frequency was normalized to the maximum frequency within each feature class to compensate for the different class sizes.

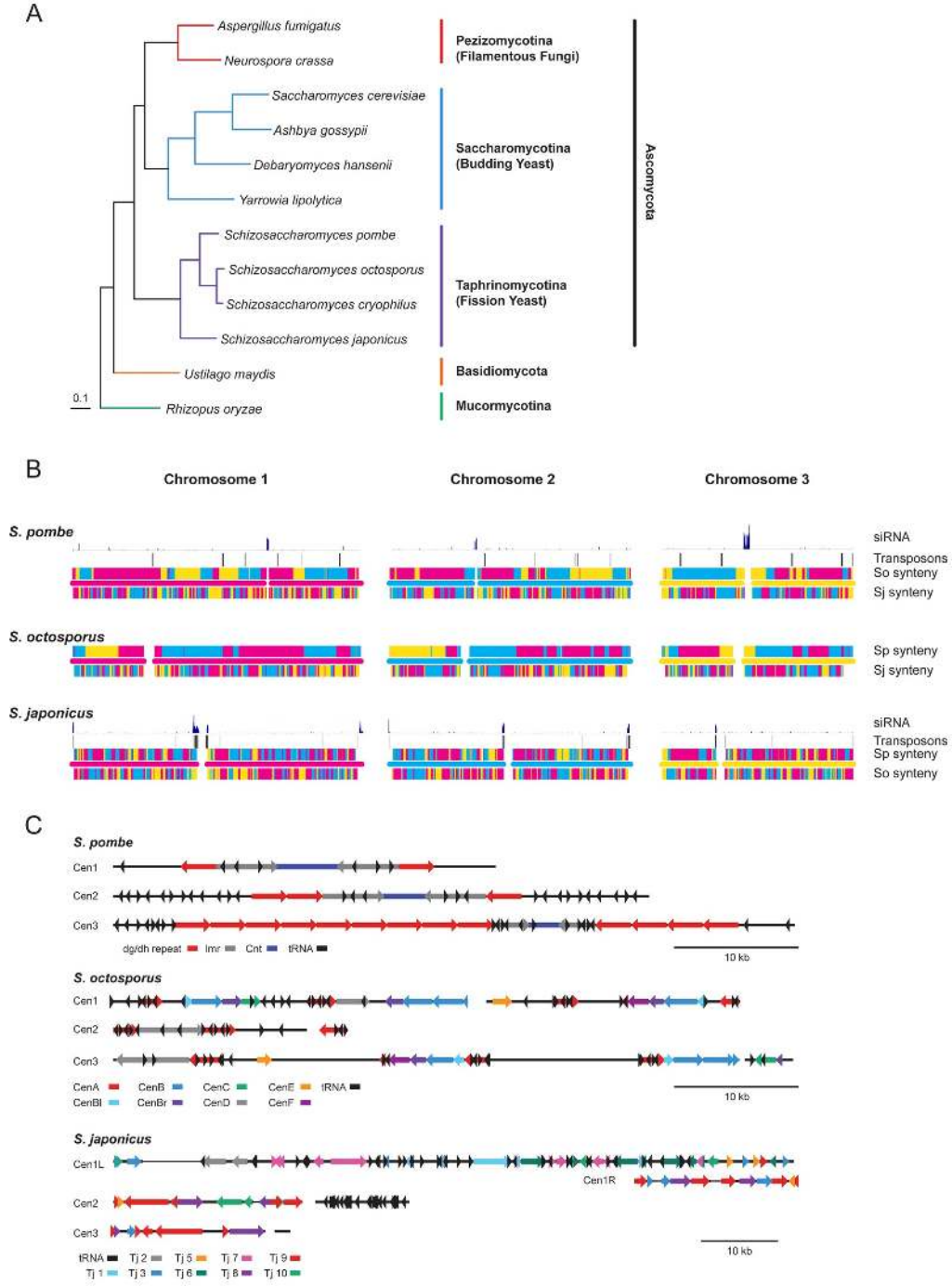
Fig. 3. Meiotic genes are subject to antisense transcription. (**A**) Examples of antisense transcription of meiotic genes. Above and below the chromosome coordinates are the coding sequence annotations on the top and bottom strand, respectively. Above and below these are the strand-specific RNA-Seq read densities on a 0-300 scale; signal above 300 is truncated to make the low amplitude signal visible. (**B**) Enrichment of GO annotations within the set of protein-coding genes with more antisense than sense transcription. All terms with a p value of less than .01 are included, except for high-level terms (i.e. biological process and molecular function).

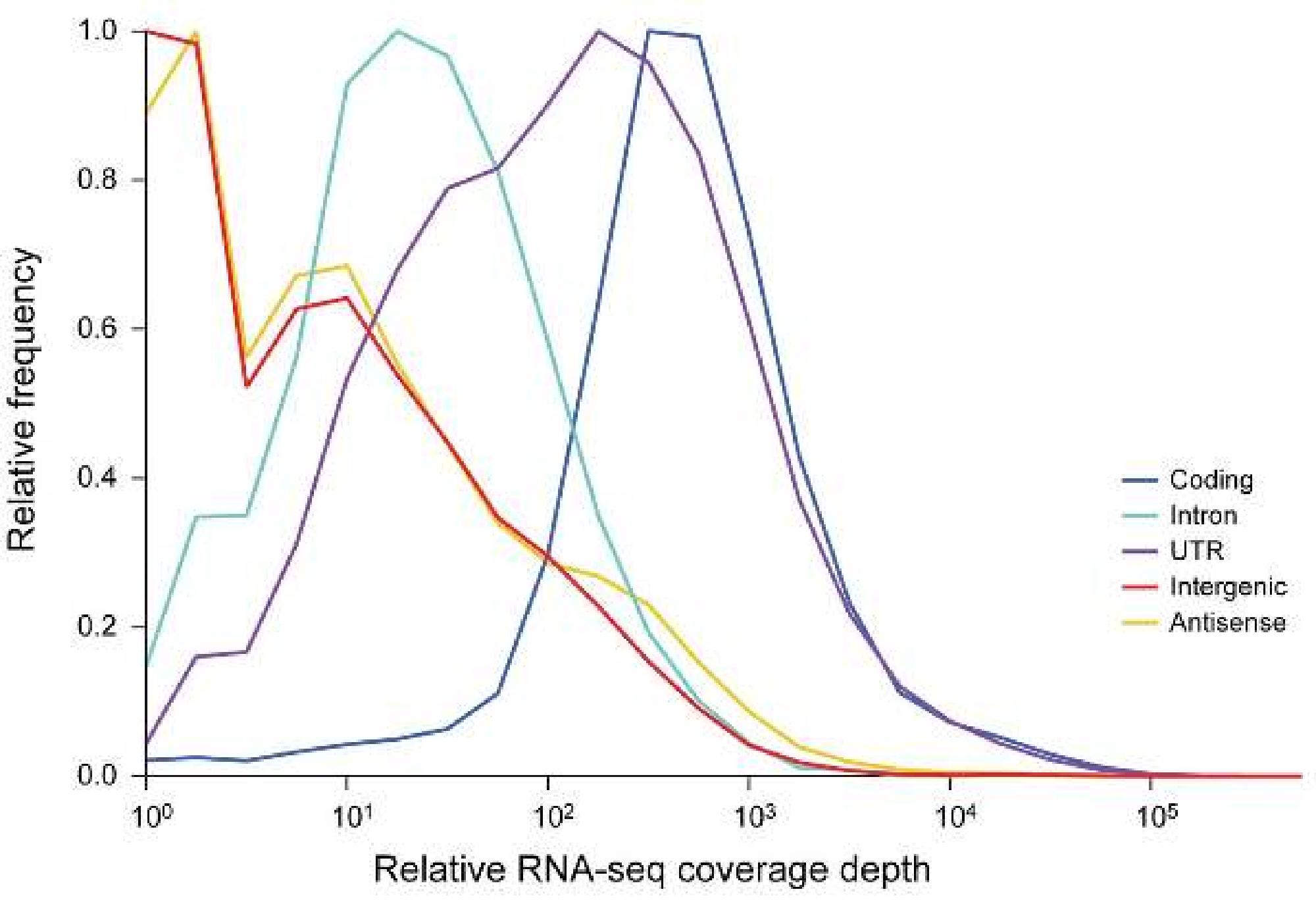
Fig. 4. Expression profiles cluster into similar patterns with conserved biological functions. (**A**) Expression clusters for each species. Gene expression profiles for each species were clustered (*4*). The size of each heat map is proportional to the number of genes in the cluster and the number of genes in each is indicated. Similar cluster sizes and patterns reflect similar expression patterns between the species. The heat-shock transcription profile is similar to log-phase growth because the transcriptional response on the 15-minute timescale used here is limited to a relatively small number of genes. (**B**) A selection of enriched GO terms for each cluster. The color intensity is proportional to the negative logarithm of the hyper-geometric p-value enrichment on a continuous scale of 0-10. Complete GO term enrichments are shown in table S26.

Fig. 5. Conserved regulatory motifs have acquired new functions and new target genes. (**A**) The enrichment of gene functional modules regulated by the Rtg3-binding motif in 23 *Ascomycota*. This motif is enriched upstream of amino acid metabolism genes in all Ascomycota. However, in fission yeast, it is specifically enriched upstream of stress-response genes. *S. cerevisiae* (*Scer*), *S. paradoxus* (*Spar*), *S. mikatae* (*Smik*), *S. bayanus* (*Sbay*), *C. glabrata* (*Cgla*), *S. castellii* (*Scas*), *K. waltii* (*Kwal*), *A. gossypii* (*Agos*), *K. lactis* (*Klac*), *S. kluyveri* (*Sklu*), *D. hansenii* (*Dhan*), *C. guilliermondii* (*Cgui*), *C. lusitaniae* (*Clus*), *C. albicans* (*Calb*), *C. tropicalis* (*Ctro*), *C. parapsilosis* (*Cpar*), *C. elongosporus* (*Celo*), *Y. lipolytica* (*Ylip*), *N. crassa* (*Ncra*), *A. nidulans* (*Anid*), *S. japonicus* (*Sjap*), *S. octosporus* (*Soct*), *S. pombe* (*Spom*). (**B**) Enrichment of Rtg3- and Aft1-binding sites in the promoters of stress response genes. Each row represents a gene. The strength of the strongest regulatory site upstream of the gene is indicated in the blue heat map. The expression of the gene in glucose depletion (gd) and early-stationery phase(es) relative to log phase is indicated in the blue-yellow heat map. Genes specific to the fission yeast clade are indicated in orange. (**C**) Enrichment of Fkh2/Mei4- and MBF-binding sites in front of antisense-transcribed genes. As in (B), but each row represents a gene with greater antisense than sense transcription. Gene associated with meiosis (*44*) are indicated in magenta.

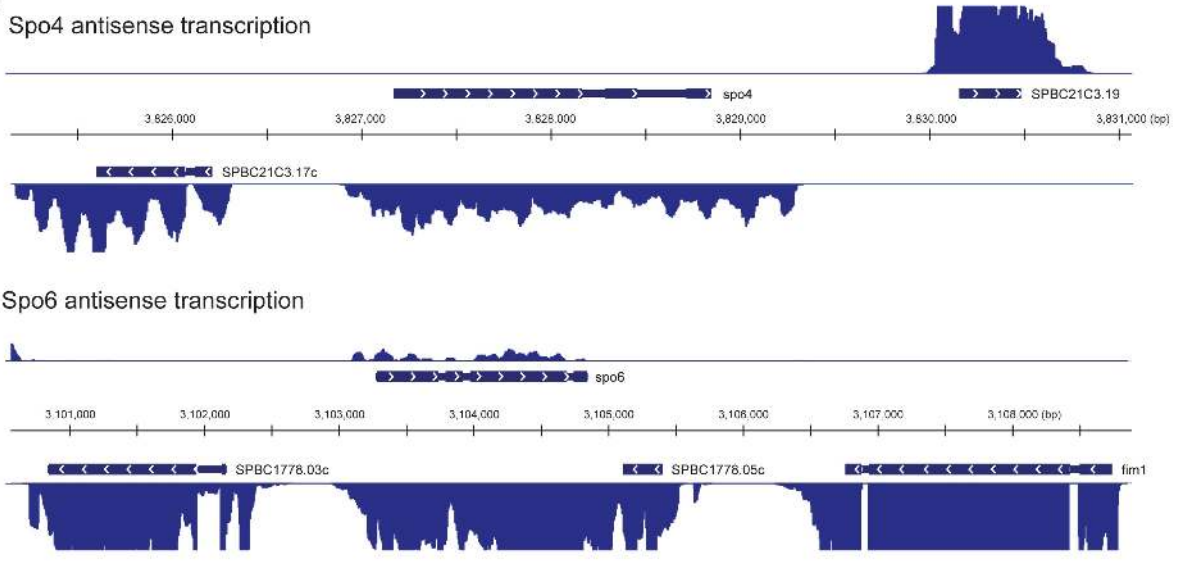
Table 1. Conservation of gene content and structure.

	Orthologous groups				Introns		
	Same	Gain	Loss	Dup	Same	Gain	Loss
S. pombe	4218	321	83	23	2901	297	27
S. octosporus	4218	133	48	5	2901	25	8
S. cryophilus	4218	283	73	11	2901	75	4
Ancestor of *Soct* and *Scry*	4218	103	44	15	2901	396	0
Ancestor of *Spom*, *Soct*, and *Scry*	4218	339	159	29	2901	415	412
S. japonicus	4218	242	0	18	2901	708	214
Ancestor of *Schizosaccharomyces*		640	745				
Total		2061	1152	101		1916	665


A
Aspergillus fumigatus
Neurospora crassa
Saccharomyces cerevisiae
Ashbya gossypii
Debaryomyces hansenii
Yarrowia lipolytica
Schizosaccharomyces pombe
Schizosaccharomyces octosporus
Schizosaccharomyces cryophilus
Schizosaccharomyces japonicus
Ustilago maydis
Rhizopus oryzae
0.1
Pezizomycotina
(Filamentous Fungi)
Saccharomycotina
(Budding Yeast)
Taphrinomycotina
(Fission Yeast)
Ascomycota
Basidiomycota
Mucormycotina
B
Chromosome 1
Chromosome 2
Chromosome 3
S. pombe
S. octosporus
S. japonicus
siRNA
Transposons
So synteny
Sj synteny
Sp synteny
C
S. pombe
Cen1
Cen2
Cen3
dg/dh repeat
Imr
Cnt
tRNA
10 kb
S. octosporus
CenA
CenB
CenC
CenE
tRNA
CenBl
CenBr
CenD
CenF
S. japonicus
Cen1L
Cen1R
tRNA
Tj 2
Tj 5
Tj 7
Tj 9
Tj 1
Tj 3
Tj 6
Tj 8
Tj 10

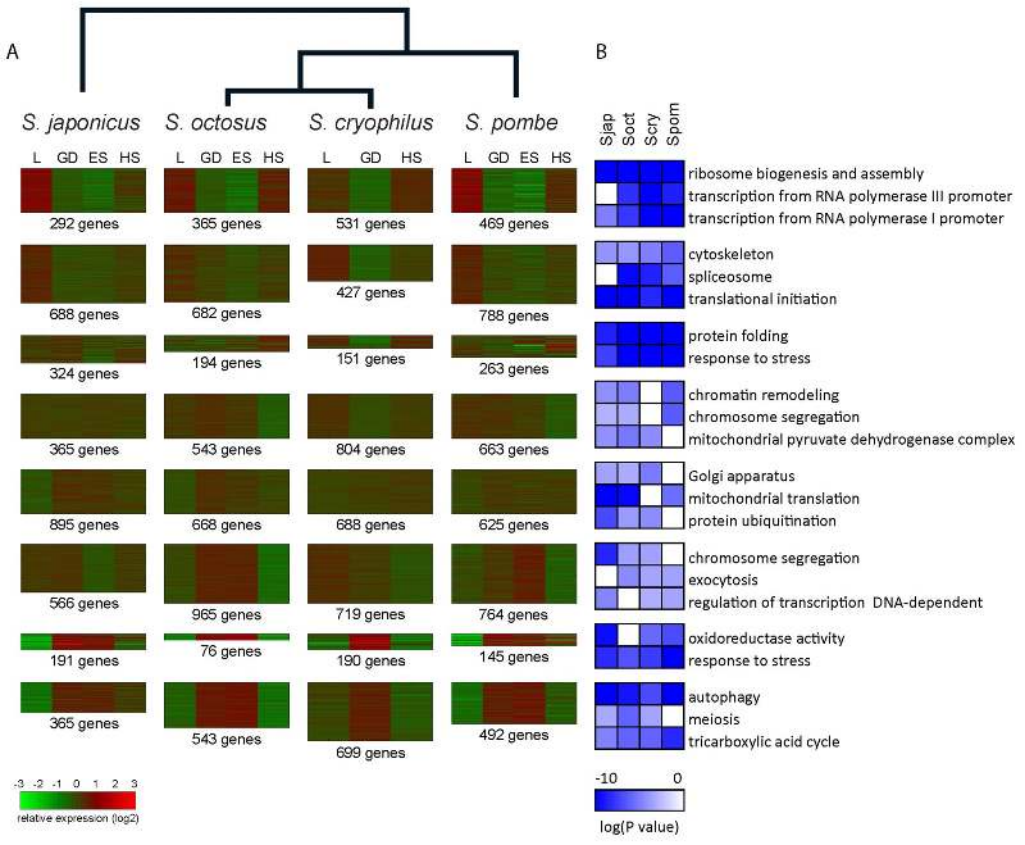

1.0
0.8
0.6
0.4
0.2
0.0
Relative frequency
10^0
10^1
10^2
10^3
10^4
10^5
Relative RNA-seq coverage depth
Coding
Intron
UTR
Intergenic
Antisense

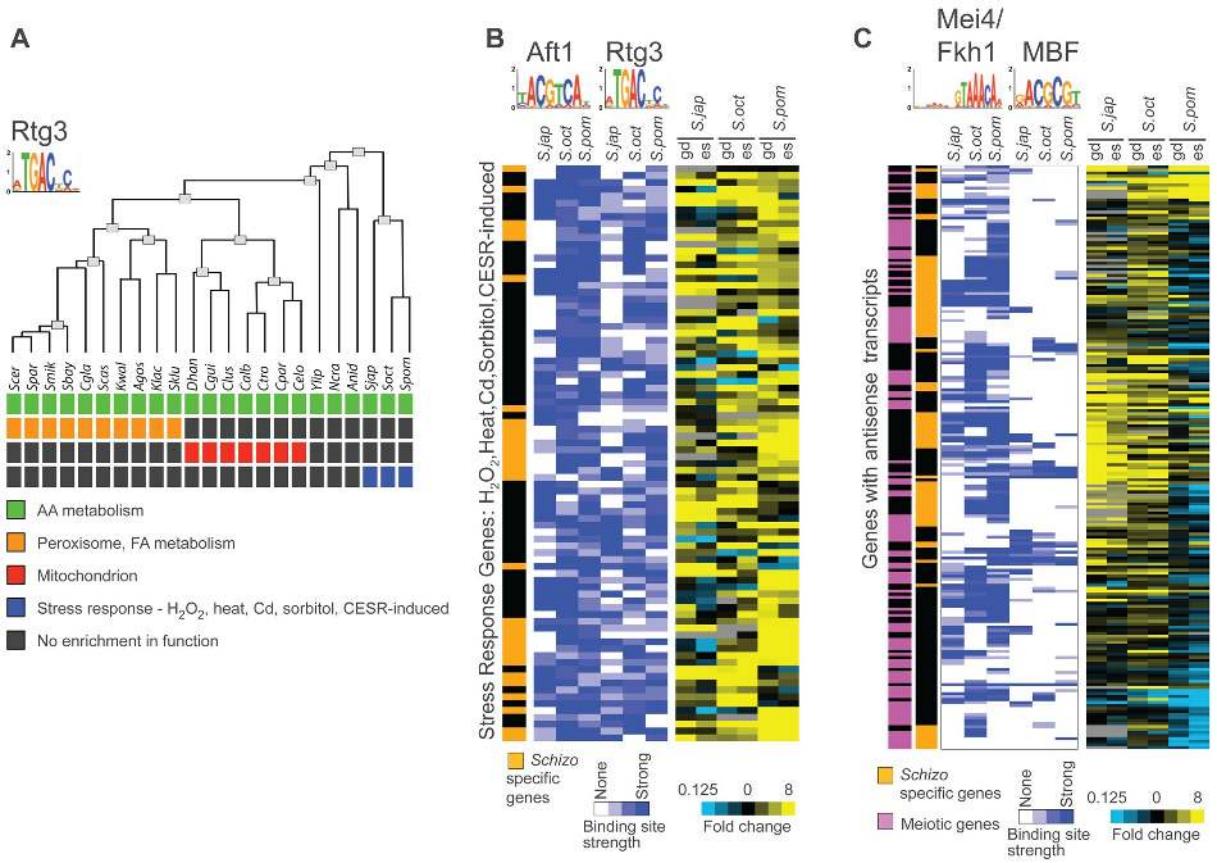
A



B

S. pombe GO annotation	p value
Meiosis	1.33×10^{-10}
Meiotic chromosome segregation	1.53×10^{-8}
Meiotic recombination	4.38×10^{-6}
Nuclear chromosome	4.54×10^{-3}
Ascospore formation	5.60×10^{-3}


A
S. japonicus
S. octosus
S. cryophilus
S. pombe
L GD ES HS
L GD ES HS
L GD HS
L GD ES HS
292 genes
365 genes
531 genes
469 genes
688 genes
682 genes
427 genes
788 genes
324 genes
194 genes
151 genes
263 genes
365 genes
543 genes
804 genes
663 genes
895 genes
668 genes
688 genes
625 genes
566 genes
965 genes
719 genes
764 genes
191 genes
76 genes
190 genes
145 genes
365 genes
543 genes
699 genes
492 genes
-3 -2 -1 0 1 2 3
relative expression (log2)
B
Sjap Soct Scry Spom
ribosome biogenesis and assembly
transcription from RNA polymerase III promoter
transcription from RNA polymerase I promoter
cytoskeleton
spliceosome
translational initiation
protein folding
response to stress
chromatin remodeling
chromosome segregation
mitochondrial pyruvate dehydrogenase complex
Golgi apparatus
mitochondrial translation
protein ubiquitination
chromosome segregation
exocytosis
regulation of transcription DNA-dependent
oxidoreductase activity
response to stress
autophagy
meiosis
tricarboxylic acid cycle
-10 0
log(P value)


A
Rtg3
Scer
Spar
Smik
Sbay
Cgla
Scas
Kwal
Agos
Klac
Sklu
Dhan
Cgui
Clus
Calb
Ctro
Cpar
Celo
Ylip
Ncra
Anid
Sjap
Soct
Spom
AA metabolism
Peroxisome, FA metabolism
Mitochondrion
Stress response - H2O2, heat, Cd, sorbitol, CESR-induced
No enrichment in function
B
Aft1
Rtg3
S.jap
S.oct
S.pom
gd
es
Stress Response Genes: H2O2,Heat,Cd,Sorbitol,CESR-induced
Schizo specific genes
None
Strong
Binding site strength
0.125
0
8
Fold change
C
Mei4/ Fkh1
MBF
Genes with antisense transcripts
Schizo specific genes
Meiotic genes