Correlations Between Gene Expression and Gene Conservation in Fission Yeast

Juan Mata and Jürg Bähler[1]

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Genes can be expressed at a wide range of levels, and they show different degrees of cross-species conservation. We compared gene expression levels to gene conservation by integrating microarray data from fission yeast (*Schizosaccharomyces pombe*) with lists of "core" genes (present in worm and budding and fission yeasts), "yeast-specific" genes (present in budding and fission yeasts, but not in worm), and "*pombe*-specific" genes (present in fission yeast only). Whereas a disproportionate number of core genes are highly expressed in vegetatively growing cells, many *pombe*-specific genes are expressed at lower levels. This bias is less pronounced in cells undergoing sexual development, when many *pombe*-specific genes become highly expressed. This implies that organism-specific proteins are more likely to function during specialized processes such as cellular differentiation. Accordingly, *pombe*-specific genes were overrepresented among genes induced during sexual development; they were particularly enriched in a group of genes induced during meiotic prophase, when homologous chromosomes pair and recombine. This raises the possibility that organism-specific genes with functions in meiotic prophase favor speciation by preventing fruitful meiosis between closely related organisms. Finally, the set of genes induced late during sexual differentiation, at the time of spore formation, was enriched in yeast-specific genes, indicating that these genes play specialized roles in ascospore development.

Comparisons of complete genome sequences provide a framework to understand the origins and evolution of eukaryotic genes. A large set of proteins is conserved in all eukaryotes, from yeast to humans. In addition, each organism contains taxon- and organism-specific genes (Chervitz et al. 1998; Rubin et al. 2000; Wood et al. 2002). It has been proposed that conserved proteins may carry out "core" cellular functions, whereas organism-specific genes may perform tasks unique to each organism (Chervitz et al. 1998; Braun et al. 2000). Conserved genes are also more likely to be essential than organism-specific genes (Fraser et al. 2000; Gönczy et al. 2000; Decottignies et al. 2003).

The fission yeast *Schizosaccharomyces pombe* has recently been sequenced; its genome is comprehensively annotated and contains ~5000 genes (Wood and Bähler 2002; Wood et al. 2002). Of these, about two-thirds have homologs in both the budding yeast *Saccharomyces cerevisiae* and the worm *Caenorhabditis elegans*. This represents a set of "core" eukaryotic genes. About 16% of *S. pombe* genes have homologs in *S. cerevisiae* but not in *C. elegans*, potentially representing "yeast-specific" genes. Finally, 14% of *S. pombe* genes do not have detected homologs in either *S. cerevisiae* or in *C. elegans*. These genes ("*pombe*-specific" genes) are likely to encode proteins specific to *S. pombe* or closely related species.

Genes with no identifiable orthologs in any other genome are called orphans (Oliver 1996). If many orphans have organism-specific functions, one would expect them to be preferentially expressed during specialized processes. To test this possibility, we have used data from genome-wide expression studies (Mata et al. 2002) to compare the expression of "core," "yeast-specific," and "*pombe*-specific" genes in fission yeast cells growing vegetatively and undergoing sexual differentiation. Sexual differentiation is a specialized developmental process that is trig-

gered by nitrogen starvation and culminates in meiotic divisions and formation of spores (Yamamoto et al. 1997). This cellular specialization is accompanied by a complex gene expression program (Mata et al. 2002).

RESULTS AND DISCUSSION

Biased Expression Levels of *Pombe*-Specific and Core Genes

We first looked at gene expression levels in fission yeast cells growing vegetatively (see Methods). A large proportion of *pombe*-specific genes were expressed at low levels (23.6% in the two bins comprising the genes expressed at lowest levels, compared with 8.3% of yeast-specific and 5.5% of core genes; see Fig. 1A; Table 1). Conversely, very few *pombe*-specific genes are expressed at high levels (2% in the two bins of most highly expressed genes, compared with 8.7% of yeast-specific and 16.8% of core genes). Core genes showed the opposite trend, with few genes among the poorly expressed and more in the highly expressed groups (Fig. 1A; Table 1). This is in accordance with findings in *C. elegans*, where conserved genes tend to be more highly expressed (The *C. elegans* Sequencing Consortium 1998). The distribution of yeast-specific genes was more symmetrical than those of *pombe*-specific and core genes and did not show any clear bias with regard to expression levels (Fig. 1A).

We then studied the expression levels of the three gene categories during sexual differentiation (Mata et al. 2002). Although many core genes were still expressed at higher levels, the bias was weaker, and many *pombe*-specific genes became strongly expressed (17.2% of genes in differentiating cells belong to the two classes of most strongly expressed genes, compared with 2% in vegetative cells; Fig. 1B; Table 1). Together, these data support the idea that core genes carry out basic cellular functions and are therefore expressed in all conditions. In contrast, many *pombe*-specific genes are only weakly expressed in vegetative cells, but
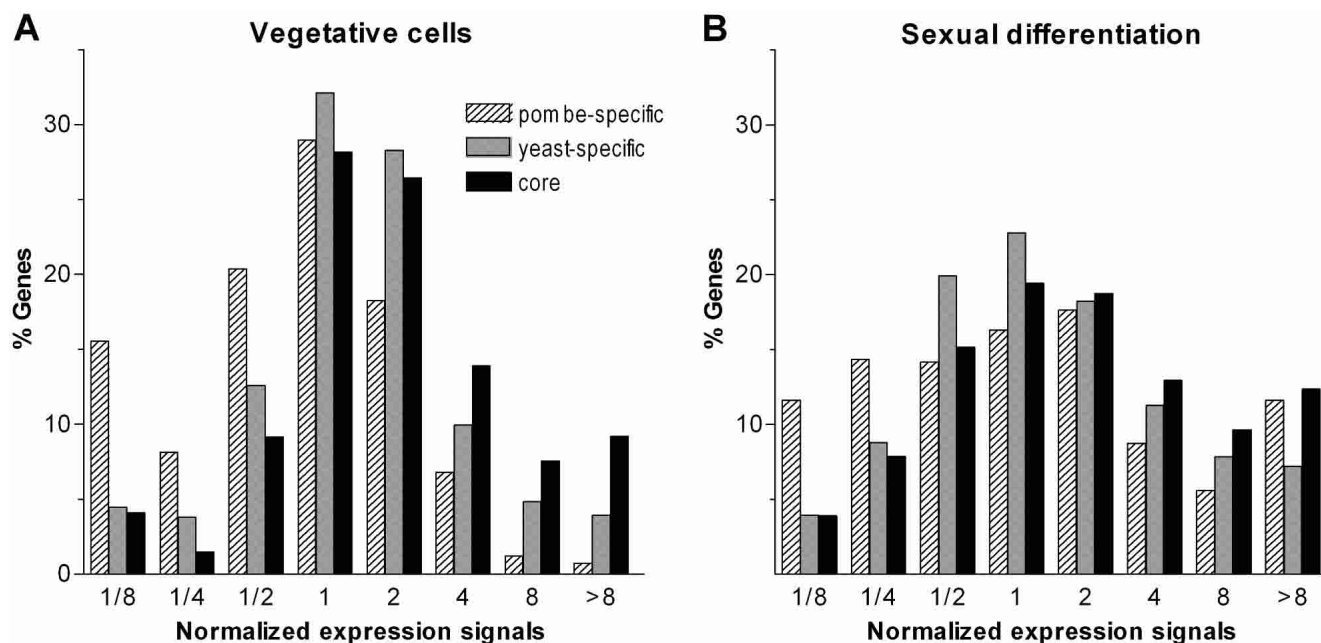
Figure 1 Expression levels in vegetative and differentiating cells. Expression levels of *pombe*-specific, yeast-specific, and core genes (*A*) in vegetative cells and (*B*) in cells undergoing meiotic divisions 5 h after inducing sexual differentiation. Microarray signal intensities were normalized to the median signal of all genes (see Methods). Genes were distributed into eight bins according to their expression signals. The values on the *X*-axes indicate the maximum normalized signals for a given bin and the minimum signal for the next bin. For example, 8 means that genes in this bin are between fourfold and eightfold more highly expressed than the median expression signal under the given condition. Absolute gene numbers are shown in Table 1.

they are induced during the specialized processes of sexual differentiation.

Genes Expressed During Sexual Differentiation Are Enriched for *Pombe*-Specific Genes

To extend this observation, we compared the distribution of the three gene categories among genes regulated during sexual differentiation (Mata et al. 2002). Unlike the comparisons in Figure 1, the following analyses do not consider absolute levels of gene expression, but are based on relative expression levels (meiotic vs. vegetative expression). We found that genes induced during sexual differentiation were enriched for *pombe*-specific genes (Fig. 2A; Table 2). Although *pombe*-specific genes make up ~13% of the genome, they represent >23% of sexually induced genes ($P < 10^{-21}$). Moreover, *pombe*-specific genes were underrepresented among genes repressed during sexual differentiation, contributing only 5.9% of those genes ($P < 10^{-7}$). We defined "*pombe*-specific genes" as those not shared with the *S. cerevisiae* and *C. elegans* genomes. To confirm these results, we used a list of orphans that contains all fission yeast genes for which no homologs have been identified in any other organism (see Methods). This list of orphans should be more accurate, as it has been generated by careful manual curation and comparison with a larger number of genomes. The conclusions obtained with this set of genes were similar to those for *pombe*-specific genes (Fig. 2B; Table 2). Core genes, on the other hand, showed a weak opposite trend (Fig. 2D; Table 2): although they represent 64% of the genes in the genome, they contribute only 53% of up-regulated genes, but 74% of down-regulated genes. Yeast-specific genes were uniformly distributed among up- and down-regulated genes (Fig. 2C).

The enrichment of orphans and *pombe*-specific genes among

genes expressed during sexual differentiation supports the hypothesis that organism-specific genes have functions in specialized processes, such as cellular differentiation. In contrast, *pombe*-specific genes were not significantly overrepresented among cell-cycle-regulated genes (G. Rusticci and J. Bähler, unpubl.), and they were only marginally enriched among those induced as part of the core environmental stress response ($P = 0.04$; Table 2; Chen et al. 2003). This indicates that cell cycle regulation and responses to stress are more conserved functions and represent core eukaryotic processes.

Table 1. Gene Numbers at Different Expression Levels

Expresssion group	Vegetative cells			Sexual differentiation		
	Pombe-specific	Yeast-specific	Core	*Pombe*-specific	Yeast-specific	Core
1/8	103	34	133	77	30	127
1/4	54	29	48	95	67	256
1/2	135	96	298	94	152	494
1	192	245	917	108	174	632
2	121	216	861	117	139	610
4	45	76	453	58	86	421
8	8	37	246	37	60	313
>8	5	30	299	77	55	402
Total	663	763	3255	663	763	3255

Genes were grouped into eight bins based on their normalized expression signals in either vegetatively growing or sporulating cells (see Fig. 1 and Methods). Gene numbers for each category and each expression group are shown. Gene numbers depend on the cutoffs used and represent estimates.
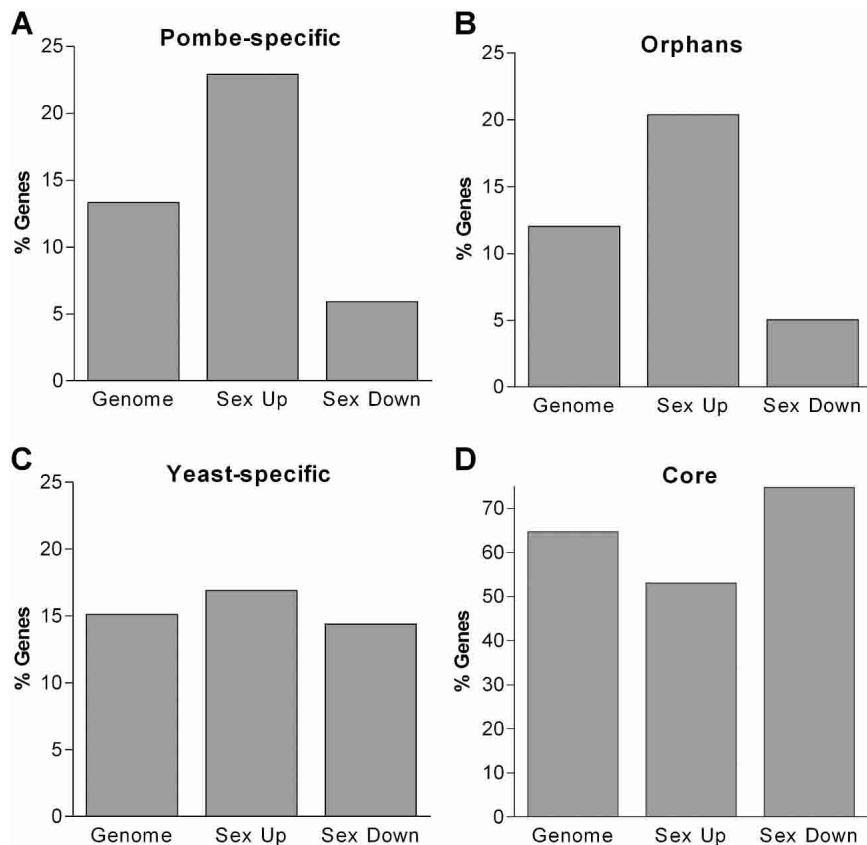
Figure 2 Distribution of conserved and lineage-specific genes during sexual differentiation. Percentages of (A) *pombe*-specific, (B) orphan, (C) yeast-specific, and (D) core genes in the fission yeast genome and among genes up-regulated (Sex Up) or down-regulated (Sex Down) during sexual differentiation. Absolute gene numbers are shown in Table 2.

otic chromosome metabolism in different organisms. Although it is difficult to directly relate sexual reproduction in multicellular organisms with sexual differentiation in unicellular eukaryotes, it is intriguing that genes involved in reproduction showed unusually high variability between human and mouse (The Mouse Genome Sequencing Consortium 2002).

Yeast-Specific Genes Are Enriched Among Genes Expressed During Spore Formation

Although yeast-specific genes were not significantly overrepresented among all genes induced during meiosis ($P > 0.05$; Fig. 2C), they were overrepresented among the late genes ($P \sim 10^{-3}$; Fig. 3C; Table 2). The expression of late genes coincides with the last steps of sexual differentiation, namely, the formation of spores. Both *S. cerevisiae* and *S. pombe* are ascomycetes, and a set of yeast-specific proteins may have conserved but specialized functions in the development of fungal ascospores.

Conclusions

We wondered whether cells use poorly and highly conserved genes in different ways. Our data show that in fission yeast, gene expression levels and transcriptional regulation correlate with gene conservation. Many conserved genes are expressed at high levels, whereas a disproportionate percentage of the poorly expressed genes are organism specific; the expression levels of the latter increase during more specialized processes such as cellular differentiation. Moreover, organism-specific genes are overrepresented among the genes induced during cellular differentiation, and they are underrepresented among those repressed. Although

Orphans Are Highly Enriched Among Genes Induced During Meiotic Prophase

The gene expression program during sexual differentiation consists of four main waves of transcription that correlate with major biological processes: nitrogen starvation-induced, early, middle, and late genes (Mata et al. 2002). *Pombe*-specific genes were overrepresented in each of the four transcriptional waves, but the early and late genes showed the highest proportions of *pombe*-specific genes (Fig. 3A; Table 2). This trend was confirmed with orphans that showed an even larger bias toward early genes, contributing >30% to these genes ($P \sim 10^{-7}$; Fig. 3B; Table 2). The great majority (84%) of these orphans are of unknown function. The early genes form a cluster that contains many genes required for chromosome pairing and recombination, indicating that some of the orphans function in these processes. This enrichment in orphans is consistent with findings that meiotic structural proteins are poorly conserved (although core recombination genes are conserved across eukaryotes; Villeneuve and Hillers 2001). Accordingly, fission yeast does not form regular synaptonemal complexes during meiotic prophase (Bähler et al. 1993), a structure that functions in meiotic chromosome metabolism in other organisms (Zickler and Kleckner 1999).

Organism-specific proteins with roles in homologous chromosome pairing during meiotic prophase may help to prevent fruitful meiosis between closely related organisms, and they may thus favor speciation. It will be interesting to see whether completely different proteins carry out conserved functions in mei-

Table 2. Gene Numbers in Different Clusters of Coexpressed Genes

Gene list	*Pombe*-specific	Orphans	Yeast-specific	Core	Total
Genome total	672	607	764	3264	5050
Sexual differentiation up	234	208	172	542	1021
−N	45	34	30	119	213
Early	31	32	16	48	101
Middle	116	106	91	327	544
Late	42	34	36	48	144
Sexual differentiation down	29	25	71	370	495
Stress: CESR up	53	46	51	189	314

Gene numbers for each category and each list of coexpressed genes are shown. Clusters of sexual differentiation genes are from Mata et al. (2002), and the corresponding percentage data are shown in Figures 2 and 3. Stress genes induced during the core environmental stress response (CESR) are from Chen et al. (2003). Gene numbers depend on the cutoffs used and represent estimates.
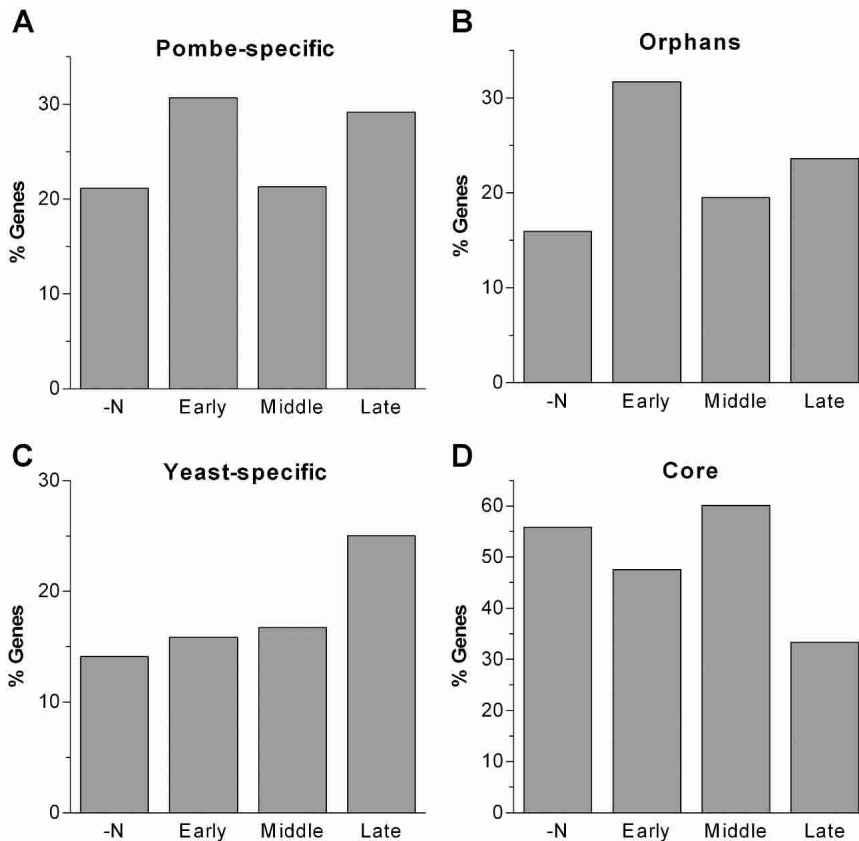
Figure 3 Distribution of conserved and lineage-specific genes at different stages of sexual differentiation. Percentages of (*A*) *pombe*-specific, (*B*) orphan, (*C*) yeast-specific, and (*D*) core genes among genes induced in response to nitrogen starvation (−N), genes induced during meiotic prophase (Early), the meiotic divisions (Middle), and the formation of spores (Late). Absolute gene numbers are shown in Table 2.

Normalization of Gene Expression Signal Intensities

To estimate absolute gene expression levels, we used fluorescence signal intensities from DNA microarray spots hybridized with vegetative and meiotic samples (Mata et al. 2002). As signal intensities may vary substantially from array to array or between different regions within one array, we normalized the signals by dividing the signal intensity of each spot by the median signal of the 900 spots surrounding it. Other factors such as the lengths and intragenic positions of microarray probes also affect signal intensities. However, these effects led to less than twofold changes in signals with our probes (Lyne et al. 2003), whereas we can obtain reliable measurements over a dynamic range of ~800-fold between the lowest and highest signals. Although DNA microarrays are normally used to measure relative gene expression levels, the signal intensities do contain valuable information on absolute gene expression levels. We used signal intensities as an approximate measure to estimate levels of gene expression.

Statistical Analyses

The statistical significance of the relative enrichment of each gene category was calculated using the hypergeometric distribution.

ACKNOWLEDGMENTS

it is not clear how well gene expression patterns reflect actual gene functions, the less conserved *pombe*- and yeast-specific genes are overrepresented in defined stages of sexual differentiation. This may reflect a particular need for organism-specific functions during specialized biological processes. A strong enrichment of organism-specific genes during defined processes such as meiotic prophase has evolutionary implications and may drive the separation between species.

METHODS

Data Sets

Data on genes induced during sexual differentiation and in response to stress were taken from gene expression profiling experiments (Mata et al. 2002; Chen et al. 2003; http://www.sanger.ac.uk/PostGenomics/S_pombe/). Lists of *pombe*-specific, yeast-specific, and core genes were created using BLASTP (Altschul et al. 1990) with a cutoff *E*-value of 0.001 and no low-complexity filtering, excluding genes encoded by the mitochondria and transposons (Wood et al. 2002). A curated list of fission yeast orphan genes was obtained from *S. pombe* GeneDB (http://www.genedb.org/genedb/pombe/index.jsp). This data set was compiled after manual inspection of alignments for the entire protein set on a gene-by-gene basis, taking into account experimental evidence, domain organization, and protein length (V. Wood, unpubl.).

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Bähler, J., Wyler, T., Loidl, J., and Kohli, J. 1993. Unusual nuclear structures in meiotic prophase of fission yeast: A cytological analysis. *J. Cell Biol.* **121:** 241–256.

Braun, E.L., Halpern, A.L., Nelson, M.A., and Natvig, D.O. 2000. Large-scale comparison of fungal sequence information: Mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*. *Genome Res.* **10:** 416–430.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282:** 2012–2018.

Chen, D., Toone, W.M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N., and Bähler, J. 2003. Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell* **14:** 214–229.

Chervitz, S.A., Aravind, L., Sherlock, G., Ball, C., Koonin, E., Dwight, S., Harris, M., Dolinski, K., Mohr, S., Smith, T., et al. 1998. Comparison of the complete protein sets of worm and yeast: Orthology and divergence. *Science* **282:** 2022–2028.

Decottignies, A., Sanchez-Perez, I., and Nurse, P. 2003. *Schizosaccharomyces pombe* essential genes: A pilot study. *Genome Res.* **13:** 399–406.

Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. 2000. Functional genomic analysis

of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408:** 325–330.

Gönczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S.J., Copley, R.R., Duperon, J., Oegema, J., Brehm, M., Cassin, E., et al. 2000. Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408:** 331–336.

Lyne, R., Burns, G., Mata, J., Rustici, G., Chen, D., Penkett, C.J., Langford, C., Vetrie, D., and Bähler, J. 2003. Whole-genome microarrays of fission yeast: Characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics* **4:** 27.

Mata, J., Lyne, R., Burns, G., and Bähler, J. 2002. The transcriptional program of meiosis and sporulation in fission yeast. *Nat. Genet.* **32:** 143–147.

The Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562.

Oliver, S.G. 1996. From DNA sequence to biological function. *Nature* **379:** 597–600.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., et al. 2000. Comparative genomics of the eukaryotes. *Science* **287:** 2204–2215.

Villeneuve, A. and Hillers, K. 2001. Whence meiosis? *Cell* **106:** 647–650.

Wood, V. and Bähler, J. 2002. How to get the best from fission yeast genome data. *Comp. Funct. Genom.* **3:** 282–288.

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.

Yamamoto, M., Imai, I., and Watanabe, Y. 1997. *S. pombe* mating and sporulation. In *The molecular and cellular biology of the yeast* Saccharomyces*: Life cycle and cell biology* (eds. J.R. Pringle et al.), pp. 1035–1106. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Zickler, D. and Kleckner, N. 1999. Meiotic chromosomes: Integrating structure and function. *Annu. Rev. Genet.* **33:** 603–754.

WEB SITE REFERENCES

http://www.genedb.org/genedb/pombe/index.jsp; *S. pombe* GeneDB.
http://www.sanger.ac.uk/PostGenomics/S_pombe/; Bähler laboratory Web site.