RESEARCH ARTICLE

Protein Rates of Evolution Are Predicted by Double-Strand Break Events, Independent of Crossing-over Rates

Claudia C. Weber and Laurence D. Hurst

Department of Biology and Biochemistry, University of Bath, Bath, Somerset, UK

Theory predicts that, owing to reduced Hill–Robertson interference, genomic regions with high crossing-over rates should experience more efficient selection. In *Saccharomyces cerevisiae* a negative correlation between the local recombination rate, assayed as meiotic double-strand breaks (DSBs), and the local rate of protein evolution has been considered consistent with such a model. Although DSBs are a prerequisite for crossing-over, they need not result in crossing-over. With recent high-resolution crossover data, we now return to this issue comparing two species of yeast. Strikingly, even allowing for crossover rates, both the rate of premeiotic DSBs and of noncrossover recombination events predict a gene's rate of evolution. This both questions the validity of prior analyses and strongly suggests that any correlation between crossover rates and rates of protein evolution could be owing to slow-evolving genes being prone to DSBs or a direct effect of DSBs on sequence evolution. To ask if classical theory of recombination has any relevance, we determine whether crossover rates predict rates of protein evolution, controlling for noncrossover DSB events, gene ontology (GO) class, gene expression, protein abundance, nucleotide content, and dispensability. We find that genes with high crossing-over rates have low rates of protein evolution after such control, although any correlation is weaker than that previously reported considering meiotic DSBs as a proxy. The data are consistent both with recombination enhancing the efficiency of purifying selection and, independently, with DSBs being associated with low rates of evolution.

Introduction

Why recombination occurs is mysterious. Classical theory concentrates on the effects recombination has on linkage disequilibrium (for review, see Otto and Lenormand 2002). Just as Fisher (1930) first noticed that in a nonrecombining population two adaptive mutations at different loci and in different individuals cannot both go to fixation, so too in a recombining genome, domains with a low recombination rate are domains where alleles in linkage disequilibrium can affect each other's fate. More generally, in finite populations selection at multiple sites could in theory reduce the effectiveness of natural selection, both positive and negative, between alleles in linkage disequilibrium (Hill and Robertson 1966; Felsenstein 1974). These effects can be configured as the consequences of a reduction in effective population size in domains of low recombination and are modulated by the extent of clustering of sites under weak selection, which reduces the opportunity for recombination between them (Comeron et al. 2008).

Because selection is less efficient in populations of small effective size, domains of low recombination are predicted to witness increased rates of fixation of deleterious mutations and decreased rates of fixation of advantageous mutations (Carvalho and Clark 1999; Comeron et al. 1999; Comeron and Kreitman 2000; Pal et al. 2001; Betancourt and Presgraves 2002). The rate of substitution of neutral mutations should in principle remain unaffected (Birky and Walsh 1988).

One common means to test this body of theory has been to ask whether, comparing genes in the same genome, the rate of protein evolution correlates with the local recombination rate (Betancourt and Presgraves 2002; Bachtrog 2003; Presgraves 2005; Zhang and Parsch 2005; Bullaughey et al. 2008; Haudry et al. 2008; Betancourt et al. 2009). In our study organism *Saccharomyces cerevisiae*, hereafter referred to as yeast, it has previously been reported that the rate of protein evolution is lowest in putative domains of high recombination (Pal et al. 2001; Connallon and Knowles 2007). Unlike studies in both mammals (Bullaughey et al. 2008) and fruit flies (see, e.g., Betancourt and Presgraves 2002; Presgraves 2005; Haddrill et al. 2007; Betancourt et al. 2009) that employed crossing-over data, the prior studies in yeast had the disadvantage of not having accurate assays of crossover per se, but instead relied on high-throughput assays of meiotic DSB events (Gerton et al. 2000). Although all crossover events start with a double-strand break (DSB) that is resolved and repaired (Baudat and Nicolas 1997), not all meiotic DSBs are resolved as crossovers. DSBs can alternatively lead to non–crossing-over events, associated with gene conversion. Measures of DSBs thus need not be accurate measures of crossing-over events.

We return to this issue now as recently Mancera et al. (2008) provided the first high-resolution counts of crossover and noncrossover recombination events in yeast (with a median of 78 bp between consecutive markers). Although a comparison of observed recombination distances (assembled over 40 years) and the DSB rate per kilobase showed that the two measures are roughly in agreement with each other (Poyatos and Hurst 2006), we find that crossover and noncrossover events are more loosely correlated than might have been expected (Spearman's $\rho = 0.2, P < 10^{-38}$), suggesting that nonspecified DSB rates have the potential to mislead. Might this be why studies in other taxa using crossing-over data yield results different from those seen in yeast? These prior studies have found no correlation with protein evolution rate in mammals (Bullaughey et al. 2008) and contradictory results in fruit flies (Larracuente et al. 2008), although the largest samples suggest weaker purifying selection on the nonrecombining chromosome IV (Haddrill et al. 2008), but not the gradual trend as seen in yeast.

Key words: double-strand break, crossing-over, rate of protein evolution.

E-mail: l.d.hurst@bath.ac.uk

Genome. Biol. Evol. 1:340–349.
doi:10.1093/gbe/evp033
Advance Access publication September 2, 2009

The new data permit us to ask two questions. First, we ask whether genes subject to high crossing-over rates really do evolve slower and whether reduced interference is a viable explanation. Second, we investigate whether DSB formation, independent of the linkage-randomizing effects of crossing-over, is correlated with rates of protein evolution. In addition, we have data on premeiotic DSBs that occur just prior to the initiation of meiosis. If DSB events are for any reason correlated with rates of protein evolution, these nonmeiotic rates might also be expected to correlate with rates of protein evolution.

Even if there is a correlation between the rate of protein evolution and crossing-over rate, this alone does not necessarily support the "efficiency of selection" model, as any correlate could be owing to a covariate. Indeed, many claimed correlates of protein rates of evolution in yeast are better explained by covariance with other parameters or data set biases (see, e.g., Pal et al. 2003; Bloom and Adami 2004; Batada, Hurst, et al. 2006; Batada, Reguly, et al. 2006; Batada et al. 2007). In yeast the assumption that genes in different recombination (or DSB) environments differ only in their rate of recombination is known to be false. For instance, dispensability (Pal and Hurst 2003), gene ontology (GO) terms (Mancera et al. 2008), and nucleotide content (Mancera et al. 2008) all covary with crossover rates (or DSB rates).

The best predictors of rates of yeast gene evolution are expression parameters (protein abundance, messenger RNA [mRNA] level, codon usage bias, etc.), such that highly expressed genes evolve slowly (Pal et al. 2001; Marais et al. 2004; Drummond et al. 2006; Wolf et al. 2008). Importantly, expression rates, averaged across many conditions and times, are also known to correlate positively with meiotic DSB rates (Pal et al. 2001), and, we can confirm, with crossover rates ($\rho = 0.084$, $P < 0.0001$). At higher resolution, crossover hot spots are enriched for genes whose expression peaks 2 h after induction of meiosis and genes whose expression decreases after 8–10 h (Mancera et al. 2008). Average meiotic expression levels of genes in DSB hot spots are elevated compared with the rest of the genome (Blitzblau et al. 2007). That DSBs in yeast are primarily found in intergenic regions containing transcription promoters (Wu and Lichten 1994; Baudat and Nicolas 1997; Gerton et al. 2000) only reinforces this connection. Similarly, histone H3 lysine 4 trimethylation is associated with both the beginning of transcribed portions of genes and site of initiation of meiotic recombination (Pokholok et al. 2005; Borde et al. 2009).

Regardless of the underlying mechanism, a negative correlation between recombination rates and protein rates is expected, even if selection efficiency is not modulated by recombination. Past work in yeast has shown that transcriptional frequency does indeed account for part of the correlation between meiotic DSB rates and nonsynonymous substitutions (Pal et al. 2001; Connallon and Knowles 2007).

We thus ask whether any correlation between rate of protein evolution and local crossover rate is robust to control for GO class, gene expression (measured three ways), nucleotide content, dispensability, premeiotic DSB rates, and meiotic noncrossover recombination rates. As both connectivity in protein-protein interaction networks and differences in hub behavior correlate with rate of evolution for artifactual reasons (Batada, Hurst, et al. 2006; Batada, Reguly, et al. 2006; Batada et al. 2007), we shall not consider them here. We also ask whether DSB events not associated with crossing-over predict rates of evolution, independent of the effects of crossing-over.

Materials and Methods

Protein Divergence

Nondubious open reading frames (ORFs) for *S. cerevisiae* and *Saccharomyces mikatae* were downloaded from the Saccharomyces Genome Database (SGD). ORF sequences with incorrect start and stop codons, premature stops within the sequence, non-A, T, G and C nucleotides, or lengths that were not multiples of three nucleotides were excluded from the set. A list of orthologs was obtained from ftp://genome-ftp.stanford.edu/pub/yeast/data_download/sequence/fungal_genomes/S_mikatae/other/MIT_mikatae_hits.txt. The DNA sequences of each ortholog pair were translated to amino acids, aligned using MUSCLE (Edgar, 2004), and reconverted to nucleotides. The nonsynonymous substitution rate (K_A) was then calculated from the nucleotide alignments according to Li (1993) with multiple hit correction according to Kimura's two-parameter model (1980). After removing the sequence for which K_A could not be calculated, 4,230 orthologs remained.

Rates of DSB events and crossovers in *S. cerevisiae* may not reflect those occurring in *S. mikatae*, for which no information is available. Therefore, an additional measure of divergence was derived from inferred *cerevisiae–paradoxus* ancestral sequences using *S. mikatae* as an outgroup. The results obtained using these data and the *cerevisiae–mikatae* distances did not differ importantly, with one exception, where a correlation moves from weakly significant to nonsignificant. This we note in the relevant section. We present in the main text the comparison with *S. mikatae*. For the parallel comparison with the *Saccharomyces paradoxus* ancestor see Supplementary Material online.

Recombination Rates

Crossover and Noncrossover Rates

Genomic coordinates for all the nondubious ORFs in the SGD data set were extracted and checked for overlaps with the positions of the intermarker intervals in the recombination data set (Mancera et al. 2008). The range of each ORF was extended upstream by 500 bp to include the promoter and 5′ untranslated regions. If overlaps were present, the mean adjusted number of crossover events in all intervals overlapping the ORF was calculated to obtain the mean crossover count across the ORF. For ORFs on the complement strand, the coordinates were first sorted in ascending order, to avoid misclassifying overlaps and nonoverlaps for complementary strand genes. To ensure that SGD ORFs had been assigned the correct coordinates, the adjusted crossover counts from the Mancera et al. (2008) data and the mean adjusted crossover counts for the ORFs were, respectively, plotted against the means of their start and end

coordinates. The peaks for the adjusted mean ORF counts match the adjusted crossover event counts well (see supplementary fig. 1, Supplementary Material online). Noncrossover recombination events, which were also provided, were assigned to ORFs in the same manner.

DSB Rates

Three different analyses have estimated DSB rates by assaying three different proteins associated with DSB formation (Spo11, Dmc1Δ, and Mre11). Spo11 DSB data (Gerton et al. 2000) are those employed by prior analyses. The mean DSB to total DNA ratio for each gene was calculated. Additionally, denoised Dmc1Δ DSB data from Buhler et al. (2007) were used to estimate the DSB rates in ORFs. The Dmc1Δ enzyme is required for meiotic DSB break repair and reveals DSBs in regions thought to be cold spots based on previous analyses. It reveals a more homogeneous distribution of hot spots and is more consistent with known recombination maps of wild-type cells, as well as crossover rates from Mancera et al. (2008).

Mre11 data provide a further measure of DSB rates. The Mre11p complex is required for DSB repair by homologous recombination. There are two time points in the data set: breaks that occur at 0 h, during G1 phase, preceding meiosis and the initiation of recombination, and breaks present at 6 h after recombination (Borde et al. 2004). The means for each gene were calculated, treating the two time points separately. This allows us to distinguish between the background DSB rate and DSBs associated with meiosis.

Estimators of Expression

Translation rate, which has a strong effect on protein evolution, but cannot be measured directly, is a correlate of transcription, protein abundance, and codon usage bias (Drummond et al. 2006). We hence use three assays to approximate translation rate.

Protein Abundance

de Godoy et al. (2008) quantified protein abundance in *S. cerevisiae* grown on Yeast Extract Peptone Dextrose (YEPD) using stable isotope labeling by amino acids in cell culture.

Codon Adaptation Index

Codon adaptation index (CAI; Sharp and Li 1987) also serves as a proxy for expression. The CAI for each *S. cerevisiae* sequence was obtained using CodonW (http://codonw.sourceforge.net) using *S. cerevisiae* reference values.

Transcriptional Frequency

If transcription and recombination are related due to chromatin state, this ought to be seen at this level, but not necessarily abundance or CAI. Therefore, the steady-state mRNA levels of each ORF were used to approximate transcriptional frequency (Holstege et al. 1998). In addition, serial analysis of gene expression data from Velculescu et al. (1997) was used to measure expression at the G1/S phase transition (as Mre11 0-h data are from G1).

Dispensability

Deutschbauer et al. (2005) classified genes as essential, slow, or non–slow growers based on their knockout phenotype on minimal medium and YEPD. The homozygous minimal medium classifications were used in this analysis to allow identification of knockouts that do no affect growth even when no copies are present and growth conditions are less favorable.

Statistics

Partial Spearman's rank correlations (hereafter referred to as β) were calculated in *R*, and *P* values for partial correlations were determined by performing 1,000 randomizations. All relationships were checked for monotonicity by eye. One exception was found, this being the relationship between transcription and crossover rates within the class of essential genes.

Results

Protein Rates of Evolution Negatively Correlate with Recombination Rates

There is a negative correlation between K_A and all four examined measures of recombination (see table 1). We can thus replicate prior results (Pal et al. 2001; Connallon and Knowles 2007) using new data. Strikingly, the Spo11 and Dmc1Δ DSB data suggest the correlation between DSB events and divergence to be stronger than that observed in the crossover data. To ask whether this might be robust, we performed a bootstrap test. Here we sample with replacement the original data for both DSB and crossover data sets. We then determine the difference in the Spearman rank correlation between paired bootstraps. We repeat this sampling process 10,000 times and ask how often among such samples the DSB bootstrap shows a more negative correlation than its paired crossover bootstrap. For both Spo11 and Dmc1Δ fewer than one in a thousand bootstraps reports the DSB correlation to be weaker than the crossover data (see supplementary table 1, Supplementary Material online). We conclude that some DSB data are a better predictor of rates of protein evolution than are crossover rates.

The above results are unchanged when we consider the divergence between *S. cerevisiae* and the ancestral sequence in the common ancestor with *S. paradoxus* (see supplementary data, Supplementary Material online) rather than K_A from the *S. cerevisiae–S. mikatae* comparison. This suggests that crossover sites moving between stable DSB hot spots cannot account for the stronger correlation between DSBs and divergence.

Functional Class Does Not Explain Reduced Evolutionary Rates in Domains of High Recombination

Mancera et al. (2008) observed a correlation between crossover hotspots and a number of GO terms. If genes

Table 1
Correlations between K_A and Crossing-over/DSBs Controlled for Codon Adaptation Index, Protein Abundance, and Transcriptional Frequency

	Spearman's ρ	Partial Correlations (β) Controlled for		
	K_A	CAI	Abundance	Transcription
Crossovers (n = 2,925)	−0.1035 $P < 10^{-7}$	−0.0919 $P < 10^{-3}$	−0.0818 $P < 10^{-3}$	−0.0756 $P < 10^{-3}$
Spo11, (n = 2,925)	−0.1745 $P < 10^{-20}$	−0.1429 $P < 10^{-3}$	−0.1333 $P < 10^{-3}$	−0.1086 $P < 10^{-3}$
Dmc1Δ (n = 2,914)	−0.18 $P < 10^{-21}$	−0.1548 $P < 10^{-3}$	−0.1461 $P < 10^{-3}$	−0.1006 $P < 10^{-3}$
Mre11 6 h (n = 2,918)	−0.0934 $P < 10^{-6}$	−0.0862 $P < 10^{-3}$	−0.0639 $P < 10^{-3}$	−0.0556 $P < 10^{-3}$
Mre11 0 h (n = 2,918)	−0.1084 $P < 10^{-8}$	−0.0895 $P < 10^{-3}$	−0.0824 $P < 10^{-3}$	−0.0704 $P < 10^{-3}$

NOTE.—Only sequences for which all three measures of expression were available were considered.

located in different recombinational environments are functionally different, could this influence overall trends in K_A?

To determine this, we performed a randomization in which genes from the same GO class were randomized with each other. In order to avoid assigning functional categories that only contain a few genes, GO Slim terms from http://downloads.yeastgenome.org/literature_curation/go_slim_mapping.tab were used. Genes were divided into bins according to crossover rate, with all genes with no observed adjusted crossover counts in one bin and three bins for high-, intermediate-, and low-crossover terciles. Different ontology terms are not evenly distributed across recombinational environments (χ^2 for independence $P = 0.0095$ for compartments; $P = 0.021$ for processes; $P = 0.0421$ for function).

If differences in K_A in different recombinational environments are solely due to crossover and not the kinds of genes that sit there, compiling a list of ontology terms found in high-recombination regions and selecting at random genes (i.e., sequences with random recombination rate) from those categories ought not to give a lower mean K_A than for the entire set.

To avoid drawing samples from a single recombination bin, two checks were performed: 1) Is any particular term unique to one crossover bin? 2) Is the number of genes that are associated with a term in any one bin higher than the sum of genes for that term in all the remaining bins? As this was not the case for any of the terms, a list of GO terms of genes from the high-crossover bin and a list of K_A values for random genes associated with each term were extracted. For high-crossover ORFs associated with more than one GO term, the category from which the K_A value was drawn was selected at random. This was repeated 1,000 times and done separately for each category (i.e., compartment, function, and process), resulting in three tables with a row for each high-crossover ORF and columns for the 1,000 sampled K_A values. Hence, the relative contributions of the different GO terms to the mean K_A were preserved. From the resulting data table, the mean K_A for each row was calculated.

Compared with the actual observed substitution rates, the randomized list for cellular compartments is not significantly different from the whole set of genes ($P = 0.0976$). Likewise, for cellular process the randomized median K_A is not significantly different from the whole set ($P = 0.053$). For molecular function, median K_A is significantly higher than the whole set (difference 0.004, $P = 0.0118$). Hence, GO terms do not account for low K_A in high-crossover genes.

The Correlation between Rates of Evolution and Recombination Rates Is Not Exclusively Owing to Skewed Nucleotide Composition

Sequences with high crossover rates are enriched for GC nucleotides (Mancera et al. 2008). Might differences in nucleotide composition explain the correlation between K_A and recombination rates? To address this, we generated a set of artificial sequence alignments that preserve codon usage and hence nucleotide bias. In order to do this, lists of all codons in all *S. mikatae* orthologs that are aligned to each of the 61 sense codons when that codon is in *S. cerevisiae* were generated. Then, an artificial sequence was derived by drawing a random sample from the list of *S. mikatae* codons for every triplet in each *S. cerevisiae* sequence. This generated a sequence alignment in which the *S. cerevisiae* gene is identical to that seen in *S. cerevisiae*, but aligned to a "pseudo" *S. mikatae* gene. K_A was calculated for each of these alignments as described above. This was repeated 1,060 times per gene. Scripts for the generation of Z scores from a given set of alignments are obtainable from CCW.

If K_A is predicted solely by nucleotide usage, there should be no relationship between the Z score and crossing-over, where Z is the location of an observation relative to the expected mean measured in standard deviations (SDs):

$$Z = \frac{\text{observed } K_A - \text{simulated mean } K_A}{\text{SD (simulated } K_A)}.$$

If Z is negative, the gene is evolving slower than expected, given the rate of evolution in the sample of pseudoalignments. For values of Z that are positive, the gene is evolving faster than expected. Hence, if there is a significant

Table 2
Correlations between *Z* and Crossing-over/DSBs Controlled for Codon Adaptation Index, Protein Abundance, and Transcriptional Frequency

	Spearman's ρ	Partial Correlation (β) Controlled for		
	Z	CAI	Abundance	Transcription
Crossovers (n = 2,925)	−0.0896 $P < 10^{-5}$	−0.0755 $P < 10^{-3}$	−0.0672 $P < 10^{-3}$	−0.0642 $P < 10^{-3}$
Spo11 (n = 2,925)	−0.1274 $P < 10^{-11}$	−0.0888 $P < 10^{-3}$	−0.0822 $P < 10^{-3}$	−0.0673 $P < 10^{-3}$
Dmc1Δ (n = 2,914)	−0.1407 $P < 10^{-13}$	−0.1092 $P < 10^{-3}$	−0.1033 $P < 10^{-3}$	−0.0707 $P < 10^{-3}$
Mre11 6 h (n = 2,918)	−0.0589 $P = 0.0015$	−0.0455 $P = 0.008$	−0.0268 $P = 0.0789$	−0.0239 $P = 0.0929$
Mre11 0 h (n = 2,918)	−0.1239 $P < 10^{-10}$	−0.1078 $P < 10^{-3}$	−0.1007 $P < 10^{-3}$	−0.0935 $P < 10^{-3}$

NOTE.—Only sequences for which all three measures of expression were available were considered.

negative relationship between the crossover rate and *Z*, this shows that the relationship between crossing-over and K_A does not depend exclusively on nucleotide content of the gene observed in *S. cerevisiae*. This is what we observe ($\rho = -0.0896$, $P < 0.0001$). However, there is also a significant negative relationship between the mean simulated K_A and crossing-over ($\rho = -0.0784$, $P < 10^{-6}$), indicating that divergence is also influenced by the factors that were controlled for in this test. Our *Z* measure, as expected, strongly correlates with K_A ($\rho = 0.9427$, $P < 0.0001$). Below we use *Z* as a composition-controlled measure of rates of evolution. Bootstrapping again suggests that Spo11 and Dmc1Δ DSB data are better predictors of *Z* than crossover rates (supplementary table 1, Supplementary Material online).

The Correlation between K_A and Recombination Is not Fully Explained by Gene Expression

In *S. cerevisiae*, there is a positive relationship between recombination and gene expression: Crossing-over is significantly positively correlated with CAI ($\rho = 0.0450$, $P = 0.0127$), protein abundance ($\rho = 0.0618$, $P = 0.0006$), and transcriptional frequency ($\rho = 0.084$, $P < 0.0001$). As recombination is associated with higher expression and high expression is associated with decreased K_A, partial correlation was used to control for nonindependence. However, covariation with protein abundance and CAI only explains part of the relationship between K_A and recombination (see table 1). This is also true for the correlation between *Z* and recombination, with the exception of meiotic (but not premeiotic) Mre11 breaks, which can be accounted for by transcriptional frequency and protein abundance (see table 2).

Abundance and transcription data exist for 3,154 and 3,676 sequences, respectively, out of a total of 4,230 orthologs. Hence, our expression data are incomplete, possibly due to low or highly conditional expression. Therefore, any control for expression based on the available information comes with a minor caveat.

These results also imply that, because controlling for CAI, protein abundance, and transcriptional frequency yields similar results, the use of codon bias as a measure of adaptation (see, e.g., Haddrill et al. 2007) must be treated with caution. Without adequate control for gene expression, enhanced adaptation cannot be inferred from an increase in the number of optimal codons.

The Correlation between K_A and Recombination Rate Is Not Entirely an Artifact of Dispensability

Different essentiality classes are not evenly distributed along the genome. Clusters of essential genes are known to reside in low-recombination regions (Pal and Hurst 2003). A priori it is hard to see why this might result in slow-evolving genes being in domains of high recombination. Nonetheless, as opportunity for positive selection might vary with dispensability, control for knockout phenotype is desirable. Similarly, Connallon and Knowles (2007) suggest that opportunities for slightly deleterious substitutions might be more common for nonessential genes.

When we analyze the data in the three proscribed classes (slow growing, non–slow growing, and essential), we see within each grouping a correlation between K_A and recombination rate (see table 3 and fig. 1). This appears to be most profound for genes resulting in slow growth on knockout. Employing *Z* as our measure the effect disappears for essential genes ($\rho = -0.0522$, $P = 0.1662$), but remains robust for slow growers ($\rho = -0.1417$, $P = 0.0009$) and non–slow growers ($\rho = -0.0874$, $P = 0.0004$). The slow rate of evolution of essential genes in domains of high recombination, we conclude, is fully explained by their skewed nucleotide usage.

Might the within-class correlations of the remaining two groups in turn be explained by covariance with transcription rates? There is no significant relationship between crossing-over and transcriptional frequency in the genes associated with non–slow growth (see table 4). Accordingly, the correlation between crossing-over and K_A in these sequences ought not to be related to gene expression. Indeed, after controlling for transcriptional frequency ($\beta = -0.08$, $P < 0.001$), protein abundance ($\beta = -0.09$, $P < 0.0001$), and CAI ($\beta = -0.09$, $P < 0.001$) the correlation between K_A and crossovers remains significant. The same is true for *Z* ($\beta = -0.0751$, $P = 0.002$ for transcription, see table 3). This is reflected in the lack of correlation between transcriptional

Table 3
Partial Spearman's Correlations between K_A/Z and Crossing-over by Dispensability Class, Controlled for CAI, Protein Abundance, and Transcriptional Frequency

	Spearman's ρ	Partial Correlation (β) Controlled for		
	Raw	CAI	Abundance	Transcription
Essential (n = 706)				
Z	−0.0522	−0.0245	−0.0169	−0.0132*
	P = 0.1662	P = 0.3007	P = 0.32468	P = 0.3656
K_A	−0.0988	−0.0798	−0.0693	−0.0562*
	P = 0.0086	P = 0.015	P = 0.039	P = 0.0859
Slow (n = 542)				
Z	−0.1417	−0.1207	−0.0632	−0.0943
	P = 0.0009	P = 0.005	P = 0.0669	P = 0.019
K_A	−0.1387	−0.1171	−0.0609	−0.0825
	P = 0.0012	P = 0.004	P = 0.0829	P = 0.03
Non-slow (n = 1,630)				
Z	−0.0874	−0.0823	−0.0846	−0.0751
	P = 0.0004	P = 0.001	P = 0.002	P = 0.002
K_A	−0.0928	−0.09	−0.09	−0.08
	P = 0.0002	$P < 0.001$	$P < 0.001$	$0 < 0.001$

NOTE.—*Note that the relationship between transcription and crossing-over rates within the essential class may not be monotonic.

frequency and crossing-over for these sequences (see table 4). For genes associated with non–slow growth on knockout, we conclude that they really do evolve slower in domains of high recombination.

For genes associated with slow growth, the picture is more complex. Here we observe that correction for protein abundance removes any significance. Are the non–slow growers and slow growers truly different, or is this lack of significant correlation in the slow growers just a sample size artifact? In order to address this, random samples of n = 542 (the number of sequences in the smallest dispensability class, the slow growers) were drawn from the non–slow genes 1,000 times. Partial correlations for Z and crossing-over, and K_A and crossing-over, controlled for protein

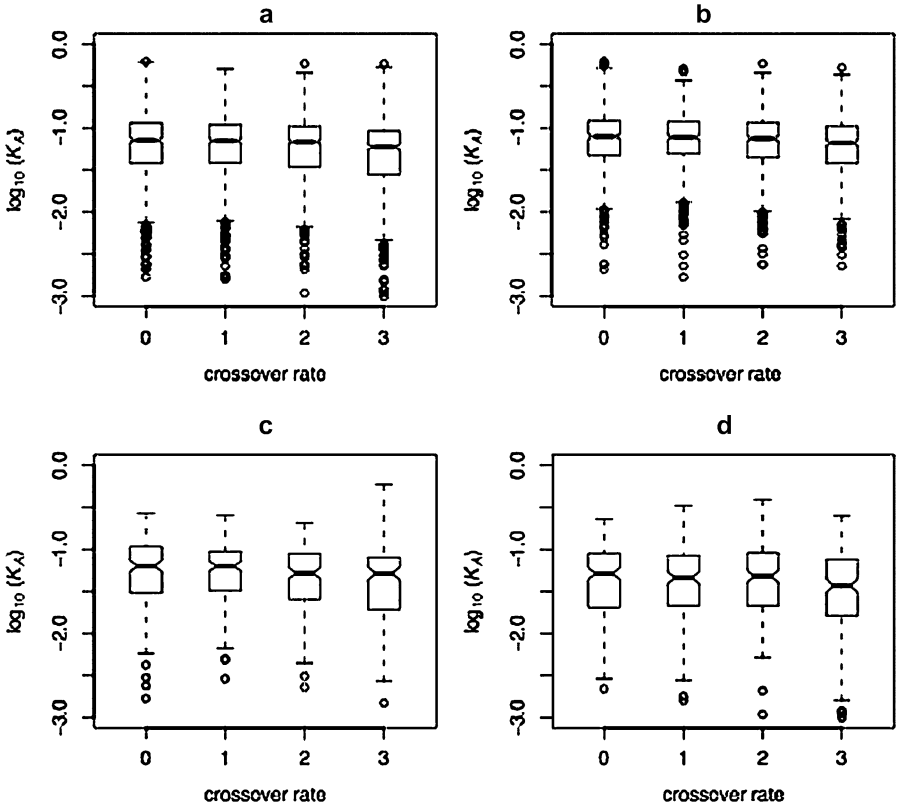

FIG. 1.—K_A versus crossover rates by knockout phenotype on minimal medium as classified by Deutschbauer et al. (2005). Genes are binned according to crossover rate, where 0 indicates no observed adjusted crossovers and 1–3 are low-, intermediate-, and high-crossover terciles. (*a*) All genes. (*b*) Knockouts with non–slow-growing phenotype. (*c*) Slow-growing knockout phenotype. (*d*) Essential genes.

Table 4
Spearman's Correlations for Transcriptional Frequency and Crossovers by Homozygous Knockout Phenotype on Minimal Medium

	All Genes (n = 2,611)	Non-slow (n = 1,598)	Slow (n = 456)	Essential (n = 557)
Spearman's ρ	0.0837, $P < 0.0001$	0.0378, $P = 0.1314$	0.1576, $P = 0.0007$	0.1364, $P = 0.0013$

NOTE.—For non–slow growers there is no significant relationship between transcription and crossovers.

abundance, were calculated as described above. From these randomizations, the probabilities of observing a partial correlation greater (i.e., less negative) than or equal to, or a P value greater than or equal to those of the slow growers were obtained. For Z, the probabilities were 0.294 for P and 0.28 for β. For K_A, the probabilities were 0.1758 for P and 0.1808 for β. We conclude that the loss of significance seen in the slow class could be a sample size effect.

Overall, we conclude that control for dispensability does not modify the conclusion that the correlation between K_A and recombination rate, as well as Z and recombination rate, is not entirely artifactual. The slow evolution of essential genes in domains of high recombination is, however, explicable in terms of skewed nucleotide usage.

Premeiotic DSB Rates Predict Rates of Protein Evolution

Although expression, nucleotide bias, and essentiality do not explain all the effect of crossing-over, our results also underscore the notion that genes subject to DSBs are not a random subset. This is strongly reinforced by consideration of rates of DSBs that occur prior to meiosis. Borde et al. (2004) looked at Mre11 deposition rates both during the meiotic period and at time zero, that is, immediately after sporulation, employing them as a pre-meiotic control. Strikingly, the correlation between K_A and premeiotic Mre11 DSBs at 0 h is stronger than that for postmeiotic Mre11 DSBs at 6 h, although the difference is not robust to bootstrapping (see supplementary table 2, Supplementary Material online), even after controlling for expression (see table 1). It is about the same as the correlation between crossover rates and K_A. When considering Z, the time zero data are more strongly correlated ($\rho = -0.1239$, $P < 0.0001$) than are the meiotic Mre11 data ($\rho = -0.0589$, $P = 0.0015$, see table 2; difference robust to bootstrapping, see supplementary table 2, Supplementary Material online).

Might the correlation between premeiotic DSB rates and rate of evolution reflect the possibility that genes prone to time zero DSBs are also prone to meiotic breaks? Although premeiotic DSB rates are correlated both with meiotic rates of DSBs (for Mre11 $\rho = 0.1829$, $P < 0.0001$) and with crossovers rates ($\rho = 0.1066$, $P < 0.0001$), controlling the Mre11 time zero data for the meiotic rates of DSB barely alters the correlation (for K_A, $\rho = -0.092$, $P < 0.001$; for Z, $\rho = -0.1147$, $P < 0.001$), suggesting that the zero time correlation is not itself owing to the fact that genes prone to premeiotic recombination are also prone to meiotic recombination. Meanwhile, control for premeiotic Mre11 breaks reduces the correlation between K_A (or Z) and meiotic Mre11 (see table 5). Indeed, the correlation between meiotic Mre11 DSBs and Z values calculated from reconstructed ancestral sequences is entirely explained by covariance with premeiotic DSBs (see supplementary table 6, Supplementary Material online).

Likewise, controlling for crossover rates does not remove the association between 0-h Mre11 breaks and K_A ($\rho = -0.0984$, $P < 0.001$) or Z ($\rho = -0.1155$, $P < 0.001$). Accordingly, the reduction in K_A associated with premeiotic DSB sites cannot be attributed to the fact that some of these sites also experience recombination. That recombination nonetheless is important is underscored by the finding that controlling the K_A (or Z) versus crossing-over correlation for the premeiotic rates remains significant (table 5).

Noncrossover Recombination Events Are Associated with Reduced Protein Divergence Rates

Any crossover event starts with a DSB but is not always resolved as a crossover event. Instead they can be resolved as noncrossover recombination events, that is, gene conversion tracts with nonreciprocal exchange. Mancera

Table 5
Spearman's Correlations between K_A/Z/Mre11 0 h and Crossing-over/DSBs and Partial Spearman's Correlations between K_A/Z/Mre11 0 h and Crossing-over/DSBs Controlled for Mre11 Breaks at 0 h (only sequences for which stable isotope labeling by amino acids in cell culture abundance and transcriptional frequency are known)

	Spearman's ρ			Partial Correlation (β) Controlled for Mre11 0 h	
	Mre11 0 h	K_A	Z	K_A	Z
Crossovers (n = 2,918)	0.1066	−0.1034	−0.0893	−0.0929	−0.0771
	$P < 0.0001$	$P < 0.0001$	$P < 0.0001$	$P < 0.001$	$P < 0.001$
Spo11 (n = 2,918)	0.1532	−0.1739	−0.127	−0.1602	−0.1101
	$P < 0.0001$	$P < 0.0001$	$P < 0.0001$	$P < 0.001$	$P < 0.001$
Dmc1Δ (n = 2,907)	0.2309	−0.1783	−0.1389	−0.1591	−0.1147
	$P < 0.0001$	$P < 0.0001$	$P < 0.0001$	$P < 0.001$	$P < 0.001$
Mre11 6 h (n = 2,918)	0.1987	−0.0934	−0.0589	−0.0738	−0.0352
	$P < 0.0001$	$P < 0.0001$	$P = 0.0015$	$P = 0.002$	$P = 0.03$

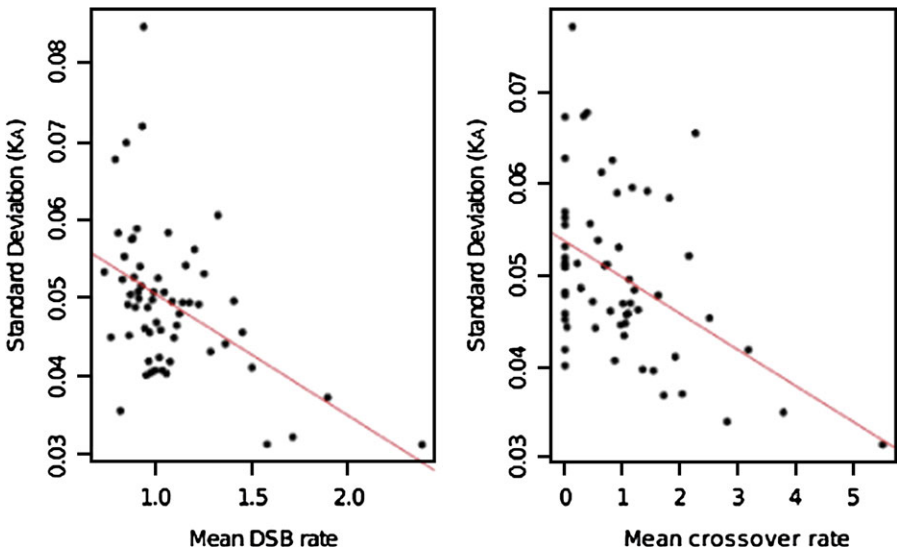

FIG. 2.—The SD of K_A decreases with increasing DSB (Gerton et al. 2000) and crossover rates (Mancera et al. 2008).

et al. (2008) estimated the rates of these. Is the rate of these events predictive of rates of protein evolution and is any such correlation itself owing to genes with frequent noncrossover events being genes with high crossover rates?

We find that, like premeiotic DSB rates, noncrossover recombination rates are inversely correlated with divergence ($\rho = -0.0629$, $P < 0.0001$ for K_A; $\rho = -0.0610$, $P < 0.0001$ for Z). We also find that crossovers and noncrossovers tend to occur at the same sites ($\rho = 0.2$, $P < 10^{-38}$). Part but not all of the effect of noncrossovers is due to covariance with crossover events ($\beta = -0.0445$, $P = 0.003$ for K_A; $\beta = -0.0448$, $P = 0.005$ for Z). As with premeiotic DSB events, this result may suggest that either some genes are prone to both being slow evolving and DSB formation or that DSBs mediate sequence evolution. However, another possibility is that noncrossovers themselves reduce linkage disequilibrium through local gene conversion effects, although the extent to which this could influence the efficacy of selection is not known.

Does the rate of noncrossover events explain why crossover rates correlate with protein evolution rates? We find that control for noncrossovers has only a modest effect on the correlation between K_A (or Z) and crossovers (for K_A, $\rho = -0.0973$, $P < 10^{-9}$ before control, $\beta = -0.0866$, $P < 0.001$ after control; for Z, $\rho = -0.0866$, $P < 10^{-7}$ before control, $\beta = -0.0761$, $P < 0.001$ after control). These results are consistent with what we observe from the premeiotic and meiotic Mre11 DSB data: Although DSBs that do not lead to crossing-over are associated with slow protein evolution, the effect of crossover events on divergence cannot be fully attributed to noncrossover DSB formation itself.

The above test is, however, not perfect. Although noncrossover recombination events are more weakly negatively correlated with divergence than are crossovers, Mancera et al. (2008) suggest that around 30% of noncrossover recombination events were likely missed due to mismatch repair or falling between markers. As this would tend to reduce the significance of the correlation observed between divergence and noncrossovers, the results for crossovers and noncrossovers cannot be directly compared.

Variance in Rates of Evolution Is Lower in Domains of High Recombination

The above covariate-controlled analyses support the thesis that crossing-over improves the efficiency of purifying selection in yeast. It is also thought that yeast is under mainly purifying selection (Connallon and Knowles 2007; Doniger et al. 2008). If both these findings are robust we might expect a further result, namely that the variance between genes in their rate of evolution should be lower in domains of high recombination. The same might be expected, however, also if the mean rate is proportional to the variance, as expected in any Poisson-like process. The test is nonetheless worth performing, not least because prior observation in Drosophila reports increased variance in K_A with increasing recombination rate (Betancourt and Presgraves 2002).

In order to test this, the data were ranked by their mean crossover rate and partitioned into 60 bins of equal size ($n = 51$). For each bin, the SD of K_A and the mean crossover rate were calculated and the Spearman's ρ between the SDs and means of the bins was calculated ($\rho = -0.3711$; see fig. 2). The K_A data were then shuffled 10,000 times to determine the probability of obtaining a value of ρ that was as extreme as or more extreme than the one observed for the ranked data ($P = 0.0012$). The same was done for the Spo11 DSB rate ($\rho = -0.4098$, $P = 0.00005$).

That there is a significant negative correlation between recombination and the SD of K_A is opposite to observations in *Drosophila*. This prior result was interpreted as support for the hypothesis that both purifying and positive selection are enhanced in high-recombination regions. Our data, by contrast, provide no prima facie evidence that domains of high recombination in yeast are commonly domains of

Table 6
Spearman's Correlations between DSBs (Borde et al., 2004) and Serial Analysis of Gene Expression Data (Velculescu et al. 1997)

	Spearman's ρ		
	G1/S	G2/M	Log
Mre11 0 h	0.08, $P < 10^{-5}$	0.02, $P = 0.2415$	0.11, $P < 10^{-9}$
Mre11 6 h	0.04, $P = 0.017$	0.02, $P = 0.1460$	0.05, $P = 0.0019$
Spo11	0.07, $P < 10^{-5}$	0.05, $P = 0.0056$	0.07, $P < 10^{-4}$

positive selection. However, it ought to be noted that the data set Betancourt and Presgraves (2002) analyzed was biased toward positively selected genes.

Discussion

From our results we conclude 1) that genes prone to DSBs tend to have low rates of protein evolution regardless of the effects of crossing-over and 2) that the thesis that crossing-over improves the efficiency of purifying selection in yeast is not falsified. The latter result is in agreement with the results from Pal et al. (2001) and Connallon and Knowles (2007) and with the view that yeast is under mainly purifying selection with little evidence for positive selection (Connallon and Knowles 2007; Doniger et al. 2008).

Our results compare well with those of Noor (2008), who found that intergenic sequence and introns in domains with high recombination rates evolve slowly in yeast. We suggest that a parsimonious explanation for all these observations is 2-fold: first that genes prone to premeiotic or meiotic DSB are, like those prone to high rates of crossing-over, those with intrinsically low rates of protein evolution; second, as argued previously, that domains of high crossing-over are domains with more efficient purifying selection. The propensity for DSB to affect conserved genes may again be mediated by expression parameters. If, for example, premeiotic DSB events are associated with transcriptional stalling (Aguilera 2002), then, a priori, we expect highly expressed genes in G1 of the cell cycle to have more premeiotic DSB events, which is indeed observed (table 6). To unify our observations with those of Noor, we need merely suppose that genes under strong constraint at the protein level are under strong constraint in the intronic and promoter regions as well.

In principle, our results might be consistent with a direct effect of DSBs (that must by necessity also be involved in crossing-over events) directly retarding sequence evolution. This would run counter to the notion that DSB sites are associated with hypermutability (Strathern et al. 1995; Yang et al. 2008). Moreover, the finding that intergenic regions in crossover hot spots have increased single nucleotide polymorphism (SNP) density (Mancera et al. 2008) argues in favor of recombination modulating interference between alleles, rather than DSB domains being regions in which new mutations are rapidly eliminated by repair processes. Moreover, with no evidence for an AT → GC bias in the SNPs in domains of high crossing-over (Noor 2008), there is no evidence that our results can be explained by biased gene conversion. That SNP rates are higher in domains of high crossing-over also argues that the correlation we see between crossover rates and protein evolution rates (and variance) is owing to more efficient purifying selection rather than an effect of DSBs during crossover events. Given that it is also quite possible that an unknown covariate exists, we conclude that we have failed to falsify the interference model rather than explicitly demonstrated it.

Supplementary Material

Supplementary data; tables 1, 2, and 6; and fig. 1 are available at *Genome Biology and Evolution* online (http://www.oxfordjournals.org/our_journals/gbe/).

Acknowledgments

C.C.W. is funded by the University of Bath. L.D.H is a Royal Society Wolfson Research Merit Award holder.

Literature Cited

Aguilera A. 2002. The connection between transcription and genomic instability. EMBO J. 21:195–201.
Bachtrog D. 2003. Protein evolution and codon usage bias on the neo-sex chromosomes of *Drosophila miranda*. Genetics. 165:1221–1232.
Batada NN, Hurst LD, Tyers M. 2006. Evolutionary and physiological importance of hub proteins. PLoS Comput Biol. 2:748–756.
Batada NN, et al. 2007. Still stratus not altocumulus: further evidence against the date/party hub distinction. PLoS Biol. 5:1202–1206.
Batada NN, et al. 2006. Stratus not altocumulus: a new view of the yeast protein interaction network. PLoS Biol. 4:1720–1731.
Baudat F, Nicolas A. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. Proc Natl Acad Sci USA. 94:5213–5218.
Betancourt AJ, Presgraves DC. 2002. Linkage limits the power of natural selection in Drosophila. Proc Natl Acad Sci USA. 99:13616–13620.
Betancourt AJ, Welch JJ, Charlesworth B. 2009. Reduced effectiveness of selection caused by a lack of recombination. Curr Biol. 19:655–660.
Birky CW Jr., Walsh JB. 1988. Effects of linkage on rates of molecular evolution. Proc Natl Acad Sci USA. 85:6414–6418.
Blitzblau HG, Bell GW, Rodriguez J, Bell SP, Hochwagen A. 2007. Mapping of meiotic single-stranded DNA reveals double-strand break hotspots near centromeres and telomeres. Curr Biol. 17:2003–2012.
Bloom JD, Adami C. 2004. Evolutionary rate depends on number of protein-protein interactions independently of gene expression level: response. BMC Evol Biol. 4:14.
Borde V, et al. 2004. Association of Mre11p with double-strand break sites during yeast meiosis. Mol Cell. 13:389–401.
Borde V, et al. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. EMBO J. 28:99–111.
Buhler C, Borde V, Lichten M. 2007. Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in *Saccharomyces cerevisiae*. PLoS Biol. 5:e324.

Bullaughey K, Przeworski M, Coop G. 2008. No effect of recombination on the efficacy of natural selection in primates. Genome Res. 18:544–554.
Carvalho AB, Clark AG. 1999. Intron size and natural selection. Nature. 401:344.
Comeron JM, Kreitman M. 2000. The correlation between intron length and recombination in Drosophila: dynamic equilibrium between mutational and selective forces. Genetics. 156:1175–1190.
Comeron JM, Kreitman M, Aguade M. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila. Genetics. 151:239–249.
Comeron JM, Williford A, Kliman RM. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity. 100:19–31.
Connallon T, Knowles LL. 2007. Recombination rate and protein evolution in yeast. BMC Evol Biol. 7:235.
de Godoy LM, et al. 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature. 455:1251–1260.
Deutschbauer AM, et al. 2005. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics. 169:1915–1925.
Doniger SW, et al. 2008. A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet. 4:e1000183.
Drummond DA, Raval A, Wilke CO. 2006. A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol. 23:327–337.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.
Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics. 78:737–756.
Fisher RA. 1930. The genetical theory of natural selection. Oxford: Clarendon Press.
Gerton JL, et al. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 97:11383–11390.
Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. Genome Biol. 8:R18.
Haddrill PR, Waldron FM, Charlesworth B. 2008. Elevated levels of expression associated with regions of the Drosophila genome that lack crossing over. Biol Lett. 4:758–761.
Haudry A, et al. 2008. Mating system and recombination affect molecular evolution in four Triticeae species. Genet Res. 90:97–109.
Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8:269–294.
Holstege FC, et al. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. Cell. 95:717–728.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16:111–120.
Larracuente AM, et al. 2008. Evolution of protein-coding genes in Drosophila. Trends Genet. 24:114–123.
Li W-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol. 36:96–99.
Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature. 454:479–485.
Marais G, Domazet-Loso T, Tautz D, Charlesworth B. 2004. Correlated evolution of synonymous and nonsynonymous sites in Drosophila. J Mol Evol. 59:771–779.
Noor MAF. 2008. Mutagenesis from meiotic recombination is not a primary driver of sequence divergence between Saccharomyces species. Mol Biol Evol. 25:2439–2444.
Otto SP, Lenormand T. 2002. Resolving the paradox of sex and recombination. Nat Rev Genet. 3:252–261.
Pal C, Hurst LD. 2003. Evidence for co-evolution of gene order and recombination rate. Nat Genet. 33:392–395.
Pal C, Papp B, Hurst LD. 2001. Does the recombination rate affect the efficiency of purifying selection? The yeast genome provides a partial answer. Mol Biol Evol. 18:2323–2326.
Pal C, Papp B, Hurst LD. 2003. Rate of evolution and gene dispensability. Nature. 421:496–497.
Pokholok DK, et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. Cell. 122:517–527.
Poyatos JF, Hurst LD. 2006. Is optimal gene order impossible? Trends Genet. 22:420–423.
Presgraves DC. 2005. Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr Biol. 15:1651–1656.
Sharp PM, Li WH. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281–1295.
Strathern JN, Shafer BK, McGill CB. 1995. DNA-synthesis errors associated with double-strand break repair. Genetics. 140:965–972.
Velculescu VE, et al. 1997. Characterization of the yeast transcriptome. Cell. 88:243–251.
Wolf MY, Wolf YI, Koonin EV. 2008. Comparable contributions of structural-functional constraints and expression level to the rate of protein sequence evolution. Biol Direct. 3:40.
Wu TC, Lichten M. 1994. Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science. 263:515–518.
Yang Y, Sterling J, Storici F, Resnick MA, Gordenin DA. 2008. Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. PLoS Genet. 4:e1000264.
Zhang Z, Parsch J. 2005. Positive correlation between evolutionary rate and recombination rate in Drosophila genes with male-biased expression. Mol Biol Evol. 22:1945–1947.

George Zhang, Associate Editor

Accepted August 28, 2009