Integrating Genomics, Bioinformatics, and Classical Genetics to Study the Effects of Recombination on Genome Evolution

John A. Birdsell

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson

This study presents compelling evidence that recombination significantly increases the silent GC content of a genome in a selectively neutral manner, resulting in a highly significant positive correlation between recombination and ''GC3s'' in the yeast *Saccharomyces cerevisiae*. Neither selection nor mutation can explain this relationship. A highly significant GC-biased mismatch repair system is documented for the first time in any member of the Kingdom Fungi. Much of the variation in the GC3s within yeast appears to result from GC-biased gene conversion. Evidence suggests that GC-biased mismatch repair exists in numerous organisms spanning six kingdoms. This transkingdom GC mismatch repair bias may have evolved in response to a ubiquitous AT mutational bias. A significant positive correlation between recombination and GC content is found in many of these same organisms, suggesting that the processes influencing the evolution of the yeast genome may be a general phenomenon. Nonrecombining regions of the genome and nonrecombining genomes would not be subject to this type of molecular drive. It is suggested that the low GC content characteristic of many nonrecombining genomes may be the result of three processes (1) a prevailing AT mutational bias, (2) random fixation of the most common types of mutation, and (3) the absence of the GC-biased gene conversion which, in recombining organisms, permits the reversal of the most common types of mutation. A model is proposed to explain the observation that introns, intergenic regions, and pseudogenes typically have lower GC content than the silent sites of corresponding open reading frames. This model is based on the observation that the greater the heterology between two sequences, the less likely it is that recombination will occur between them. According to this ''Constraint'' hypothesis, the formation and propagation of heteroduplex DNA is expected to occur, on average, more frequently within conserved coding and regulatory regions of the genome. In organisms possessing GC-biased mismatch repair, this would enhance the GC content of these regions through biased gene conversion. These findings have a number of important implications for the way we view genome evolution and suggest a new model for the evolution of sex.

Introduction

The genomes of warm-blooded vertebrates consist of large regions (>300 kb) of relatively homogeneous GC content termed isochores (Bernardi 1986, 2000). Nonvertebrate organisms also show distinctive genomic GC compositional patterns (Nekrutenko and Li 2000) as do some plants (Matassi et al. 1989; Nekrutenko and Li 2000) and the yeast *Saccharomyces cerevisiae* (Bradnam et al. 1999). This article focuses primarily on *S. cerevisiae*; however, it appears that the observations made regarding this organism may have a general applicability to organisms spanning several kingdoms.

The hypotheses proposed to explain the heterogeneity in GC content are those that favor selection (Bernardi 1986, 2000; Charlesworth 1994), regional mutational biases (Sueoka 1962; Filipski 1987; Wolfe, Sharp, and Li 1989; Gu and Li 1994; Francino and Ochman 1999), or biased gene conversion (BGC) (Brown and Jiricny 1989; Holmquist 1992; Eyre-Walker 1993; Charlesworth 1994; Galtier et al. 2001). This study presents evidence in favor of the BGC model, according to which GC-biased mismatch repair results in GC-biased gene conversion within the heteroduplexes formed during recombination. Over an evolutionary time scale, these processes result in a positive relationship between recombination and GC content.

Positive Correlation Between Recombination and GC Content

Positive correlations have been found between recombination and GC content in humans (Ikemura and Wada 1991; Eyre-Walker 1993; Eisenbarth et al. 2000, 2001; Fullerton, Bernardo Carvalho, and Clark 2001; Galtier et al. 2001; unpublished data), birds (Hurst, Brunton, and Smith 1999; Galtier et al. 2001; unpublished data), rodents (Williams and Hurst 2000), worms (Marais, Mouchiroud, and Duret 2001), insects (Marais, Mouchiroud, and Duret 2001; Takano-Shimizu 2001), and plants (unpublished data).

Recombination and GC Content in the Yeast *S. cerevisiae*

Gerton et al. (2000) used DNA microarrays, in an elegant and pioneering study, to map the relative rate of recombination throughout the *S. cerevisiae* genome at a resolution of about 1–2 kb. This study revealed a genomewide correlation between recombination and total GC content, a relationship that had previously been observed only on chromosome III (Sharp and Lloyd 1993). When the GC content within a 5-kb window was examined, there was a total of 221 GC peaks in which the GC content was >3% higher than the chromosome mean (Gerton et al. 2000). There was a total of 177 recombination hot spots. If these were distributed at random, one would expect 18 hot spots within 2.5 kb of a peak; however, there were 99 peak-associated hot spots

Key words: *Saccharomyces cerevisiae*, recombination, GC content, biased gene conversion, GC-biased mismatch repair, evolution of isochores, evolution of sex.

Address for correspondence and reprints: John A. Birdsell, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85121. E-mail: birdsell@email.arizona.edu.

Mol. Biol. Evol. 19(7):1181–1197. 2002

($P < 0.001$) (Gerton et al. 2000). The GC content of a 5-kb window with its center at the middle of each hot spot exceeded the mean GC content of the chromosome in 162 of 177 hot spots ($P < 0.001$). A significant correlation between the ranking of the hot spots and their GC content was also found ($P < 0.002$) (Gerton et al. 2000).

Because the total GC content of a gene is not particularly sensitive to mutational or substitutional biases, an analysis designed to provide more power to determine the relationship between recombination and GC content was undertaken. This study made use of the fact that silent GC content (''GC3s'') is the most sensitive measure of mutational or substitutional biases.

Methods

Two different methods were used to analyze the relationship between recombination and GC content in the yeast genome. In the first, the correlation between recombination and GC content was determined directly using the data from Gerton et al. (2000). This analysis was carried out at three different scales. In the first, correlations were made between measures of GC content and the relative recombination rate of 6,143 yeast open reading frames (ORFs) as determined from mean array values of seven microarray experiments performed by Gerton et al. (2000). The second analysis was made by correlating the mean GC3s of 10 sets of 50 loci with the mean recombination rate of each set. Mean array values of the seven experiments were sorted according to magnitude, and groups of 50 loci having contiguous array values were taken, starting with the 50 coldest loci and proceeding to the 50 hottest loci. The third analysis made use of the yeast gene duplication database (http://acer.gen.tcd.ie/~khwolfe/yeast/), which was searched (using the Smith-Waterman algorithm [Pearson 1991]) for paralogs of recombinationally hot loci. Forty-seven sets of paralogous genes, having expected values of zero, were found. The paralogs were grouped into two categories: those that were recombinationally hot (as defined by Gerton et al. 2000) and those that were not. Most genes had only one paralog; however, a few had multiple paralogs. When there was more than one hot or cold locus within a gene family, values were pooled and means were compared.

The second major approach used in this study was to analyze mismatch repair data from 12 studies to determine whether there was any repair bias. Six of these studies involved mismatch repair of heteroduplex plasmid DNA in mitotic wild-type strains. These studies used plasmids containing a reporter gene having a defined mismatch in a defined orientation. The correction of this mismatch could give rise to one of the two visibly distinct colony phenotypes depending upon the direction of correction. A proximally located nick can confer a very slight, though sometimes significant, bias in repair to the nicked strand (i.e., the base on the nicked strand is replaced) (Yang et al. 1999). If the nick is 3.5 kb or further away from the mismatch, it does not bias mismatch repair in favor of either strand (Bishop, Andersen, and Kolodner 1989; Yang et al. 1999). In five of the studies analyzed, the nick was positioned at least 3.5 kb away from the mismatch (Kramer et al. 1989; Kunz, Kang, and Kohalmi 1991; Kang and Kunz 1992; Yang et al. 1996, 1999). In the sixth study, the plasmid was ligated such that between 70% and 95% of all plasmids were covalently closed circles (Bishop, Andersen, and Kolodner 1989). Only corrections of heteromismatches (e.g., G/T, A/C, etc.) in wild-type strains were considered. Throughout this article, the terms ''GC'' and ''AT'' are used to refer to G/C or C/G and A/T or T/A base pairs, respectively. Mismatches involving the same two nucleotides in opposite orientations (e.g., G/T and T/G) were pooled within the same experiment, but the pooling of the observations between experiments was determined to be statistically inappropriate by means of a heterogeneity chi-square test (Zar 1984, pp. 49–52). Another six studies were analyzed to determine the direction of repair of meiotic-induced heteroduplexes in the *HIS4* recombination hot spot. In this analysis, segregation ratios of 6:2, 2:6, 7:1, and 1:7 were counted as one conversion event, whereas segregation ratios of 8:0 and 0:8 were counted as two conversion events.

Global recombinational data were obtained from the data of Gerton et al. (2000) at http://derisilab.uscf.edu/hotspots/. The yeast coding and intergenic sequences were obtained from the Stanford Genome database at http://genome-www.stanford.edu/Saccharomyces/. The ORF nucleotide content was determined using the General Codon Usage Analysis package available at http://www.bioinf.org/vibe/software/gcua/download.html (McInerney 1998). The analysis of codon usage bias and intergenic and intronic GC contents was performed using the Molecular Evolutionary Analysis Package (Version 6/22/00) kindly provided by Etsuko Moriyama (emoriyama2@unlnotes.unl.edu.)

The GenBank sequences of the *S. cerevisiae* intron containing ORFs along with the sequences of their introns were kindly provided by Francis Clarke at http://www.maths.uq.edu.au/~fc/datasets/. Redundant sequences were removed using CLEANUP (Grillo et al. 1996) available at http://bighost.area.ba.cnr.it/BIG/CleanUP/. The sequences of 697 yeast regulatory elements were obtained from the Promoter Database of *Saccharomyces cerevisisae* at http://cgsigma.cshl.org/jian/.

To examine the effects of BGC within a gene that has recently changed its recombinational environment, I analyzed substitutions within the *Mus musculus Fxy* gene. Coding sequences for human (AF035360), *M. musculus* (AF026565), *M. spretus* (AF186460), and *Rattus novegicus* (AF186461) *Fxy* genes were aligned in Clustal X. The sequences were highly similar, and the alignments produced no gaps. Ancestral sequences were reconstructed using maximum likelihood analysis (codeml) no molecular clock (unrooted tree) option implemented in PAML (3.0c) (Yang 1997) (http://abacus.gene.ucl.ac.uk/software/paml.html). The number and direction of silent third position substitutions were compared with the expected number on the basis of the third

position base composition of the inferred ancestral sequence.

All chi-square calculations used the Yates correction for continuity (Zar 1984, p. 48). The Kolmogorav-Smirnov tests (Zar 1984, p. 91) were used to test for normality, and the parametric tests were used when appropriate. The arcsine transformation was performed on proportional data (Zar 1984, p. 239). The measures of variance are standard errors unless otherwise stated. Tests of significance were two tailed, except for tests of correlation for which one-tailed tests are appropriate (Zar 1984, p. 309). BLASTP searches (Altschul et al. 1997) for homologs of known GC-biased mismatch repair enzymes, such as *Escherichia coli* MutYprotein and human TDG protein, were performed at http://www.ncbi.nlm.nih.gov/BLAST/, and tBLASTn searches were performed against a series of partially completed genomes at the TIGR BLAST site http://tigrblast.tigr.org/tgi/.

Results and Discussion

Mismatch Repair Bias

To determine whether there is any evidence of a mismatch repair bias in *S. cerevisiae*, an analysis was made of the repair of heteroduplex DNA in mitotic cells. Of a total of 72,971 repaired heteromismatches, 42,242 (57.9%) were repaired to G/C or C/G, whereas only 30,729 (42.1%) were repaired to A/T or T/A (tables 1 and 2). This represents a highly significant GC bias according to a Wilcoxon test (performed on the number of mismatches corrected to GC vs. the number corrected to AT for all the 50 experiments), $Z = -5.09$ ($P = 3.6 \times 10^{-7}$). Out of the 50 experiments, 36 showed a significant repair bias toward GC, whereas only 3 showed a significant repair bias toward AT. This difference (36 to 3) is, itself, highly significant, $\chi^2 = 26.3$ ($P = 2.9 \times 10^{-7}$). The mean ratio of repair to GC versus AT for all the 50 experiments is 1.48 ± 0.11 to 1. The GC repair bias was most pronounced for G/T mismatches which exhibited a mean bias of 1.71 ± 0.338 to 1 ($n = 15$), followed by C/T mismatches (1.50 ± 0.14 to 1; $n = 9$), A/G mismatches (1.38 ± 0.13 to 1; $n = 9$), and A/C mismatches (1.31 ± 0.05 to 1; $n = 17$).

To determine whether distally located nicks (3.5 kb or further away from the mismatch) were responsible for the observed GC bias in repair, a signed rank test was performed on the number of mismatches corrected to GC versus AT for the 23 experiments having an A or a T on the nicked strand. This demonstrated a very significant GC bias, $Z = -2.71$ ($P = 0.0067$). The 23 experiments having a G or a C on the nicked strand had an even more significant GC bias, $Z = -4.1$ ($P = 2.0 \times 10^{-5}$). Although there may be a slight repair bias to the strand possessing the distally located nick, it is clear that regardless of where the nick is located, there is a highly significant GC repair bias. In addition, there was no evidence of any significant difference in the relative efficiency of repair of different heteromismatches, as determined by a single factor ANOVA, $F = 1.87$, df = 7 ($P = 0.10$).

The mismatch repair studies analyzed in the preceding section (see *Methods*) all involved the repair of plasmid DNA in mitotically dividing cells. To determine whether similar repair biases exist in chromosomal DNA in meiotic cells, six studies involving a total of 2,148 informative gene conversion events were analyzed. Of these, 1,186 were corrected to GC or CG, whereas 962 were corrected to AT or TA (table 3). A comparison of the number of mismatches corrected in each direction revealed a significant GC bias, Wilcoxon $Z = -2.04$, ($P = 0.041$). The mean bias of all 15 experiments was 1.22 ± 0.10 to 1. A comparison of the mean mitotic repair bias (1.48 ± 0.11) shows that although this bias was greater than the mean meiotic repair bias (1.22 ± 0.10), the difference was not significant (Mann-Whitney $Z = -1.76$, $P = 0.08$).

These results provide the first evidence, in any fungus, of a significant GC mismatch repair bias. This bias is found in both meiosis and mitosis and suggests that *S. cerevisiae* may possess hereto uncharacterized mismatch-specific thymine glycosylase and adenine glycosylase activities. Protein blast searches did not reveal any ORFs with significant homology to known mismatch-specific adenine or thymine glycosylases. I suggest that genes, of as yet uncharacterized function, may be responsible for the observed mismatch repair biases.

It is noteworthy that the relative repair biases associated with different mismatches in yeast (G/T > C/T > A/G > A/C) are the same as those found through the analysis of the simian mismatch repair data of Brown and Jiricny (1988) (i.e., G/T > C/T > A/G > A/C). The evolution of similar mismatch repair biases may have occurred as a response to similar underlying biological phenomena.

The fact that *S. cerevisiae* does not possess 5-methylcytosine (Proffitt et al. 1984) may be reflected in the differences between the relative repair biases of G/T mismatches in yeast and mammals. The mutagenic potential of 5-methylcytosine is well known (Coulondre et al. 1978; Duncan and Miller 1980), and mammals possess substantial quantities of 5-methylcytosine. Not surprisingly, mammalian cells have evolved very efficient mechanisms to repair G/T mismatches and show a highly significant bias in favor of GC over AT of 24 to 1 (Brown and Jiricny 1988). Compare this with the more modest 1.71 to 1 GC bias seen in yeast which lacks 5-methylcytosine. I suggest that these differences may, in part, be attributable to the amount of 5-methylcytosine within the genomes of these two types of organisms and that, in general, GC mismatch repair biases will be found to be substantially greater in aerobic organisms possessing 5-methylcytosine.

Recombination versus GC Content

Within the 6,143 yeast ORFs analyzed, there is a highly significant positive correlation between silent GC content (GC3s) and recombination (fig. 1 and table 4). The mean GC content of first and second codon positions (GC1 + GC2)/2 also shows a significant, though much lower, correlation with recombination. The 100

Table 1
Summary of the Mitotic Heteromismatch Repair Data from Six Published Studies

Reference[a]	Mismatch	Total number of colonies	Proportion Repaired	Proportion Repaired to GC	Proportion Repaired to AT	Ratio of Repair GC/AT	Base on Nicked Strand	Allele on Nicked Strand
1	GT	2,901	0.950	0.862	0.138	6.25	G	+
1	GT	3,774	0.880	0.647	0.353	1.83	G	−
1	TG	1,413	0.900	0.633	0.367	1.72	T	−
1	TG	3,461	0.910	0.600	0.400	1.50	T	+
1	CA	2,839	0.650	0.595	0.405	1.47	C	+
1	CA	1,909	0.750	0.522	0.478	1.09	C	−
1	CA	2,216	0.590	0.591	0.409	1.45	C	+
1	AC	3,233	0.750	0.617	0.383	1.61	A	−
1	AC	2,802	0.790	0.593	0.407	1.46	A	+
1	AC	3,805	0.640	0.617	0.383	1.61	A	−
2	GT	1,545	0.880	0.667	0.333	2.00	G	+
2	GT	3,894	0.850	0.643	0.357	1.80	G	−
2	TG	1,485	0.850	0.625	0.375	1.69	T	−
2	AC	1,729	0.750	0.500	0.500	1.00	A	−
2	AC	2,086	0.670	0.524	0.476	1.10	A	+
2	CA	3,110	0.590	0.588	0.412	1.43	C	+
3	TG	2,003	0.980	0.535	0.465	1.15	T	−
3	TG	1,060	0.960	0.487	0.513	0.95	T	+
3	CA	1,314	0.940	0.578	0.422	1.37	C	+
3	CA	1,308	0.960	0.600	0.400	1.50	C	−
3	CA	843	0.991	0.521	0.479	1.09	C	+
3	AC	478	0.889	0.485	0.515	0.94	A	−
3	GT	927	0.890	0.463	0.537	0.86	G	+
3	TG	800	0.990	0.495	0.505	0.98	T	−
4	AG	5,856	0.930	0.455	0.545	0.83	A	−
4	GA	3,616	0.860	0.571	0.429	1.33	G	+
4	CT	6,071	0.900	0.565	0.435	1.30	C	+
4	TC	5,587	0.920	0.483	0.517	0.94	T	−
5	GT	843	0.940	0.588	0.412	1.43	T	−
5	GT	833	0.890	0.533	0.467	1.14	T	−
5	GT	1,445	0.830	0.580	0.420	1.38	T	−
5	AC	1,016	0.910	0.546	0.454	1.20	C	+
5	AC	868	0.859	0.555	0.445	1.25	C	+
5	AC	1,861	0.830	0.578	0.422	1.37	C	+
5	AG	157	0.943	0.615	0.385	1.60	G	−
5	AG	662	0.881	0.614	0.386	1.59	G	−
5	AG	527	0.801	0.602	0.398	1.51	G	−
5	GA	185	0.908	0.423	0.577	0.73	A	−
5	GA	1,109	0.760	0.571	0.429	1.33	A	−
5	GA	402	0.649	0.617	0.383	1.61	A	−
5	TC	542	0.921	0.677	0.323	2.10	C	+
5	TC	1,678	0.800	0.647	0.353	1.83	C	+
5	TC	977	0.660	0.678	0.322	2.10	C	+
5	CT	216	0.921	0.578	0.422	1.37	T	+
5	CT	828	0.850	0.578	0.422	1.37	T	+
5	CT	596	0.659	0.545	0.455	1.20	T	+
6	AC	403	0.819	0.573	0.427	1.34	NA	NA
6	GT	171	0.795	0.500	0.500	1.00	NA	NA
6	GA	143	0.811	0.655	0.345	1.90	NA	NA
6	TC	151	0.709	0.561	0.439	1.28	NA	NA

[a] References: 1. Yang et al. 1999; 2. Yang et al. 1996; 3. Kang and Kunz 1992; 4. Kunz, Kang, and Kohalmi 1991; 5. Kramer et al. 1989; 6. Bishop, Andersen, and Kolodner. 1989. NA: not applicable

hottest loci have a significantly greater GC3s (48.7% ± 0.80%) than the 100 coldest loci (34.85% ± 0.44%), Mann-Whitney $Z = -10.554$ ($P = 4.8 \times 10^{-26}$).

The GC3s is also highly correlated with recombination rate within the 10 sets of 50 loci, (fig. 2), and increases monotonically with increasing levels of recombination for all groups, except the last group containing the 50 hottest loci. This last group actually has a significantly lower GC3s than the preceding group of 50 (47.0 ± 1.13 vs. 50.4 ± 1.10, Mann-Whitney $Z = -2.20$, $P = 0.028$). If recombination is mutagenic and does have an AT bias, then in extremely recombinagenic loci, this mutational effect may slightly overcome the substitutional bias toward GC caused by BGC (see subsequent discussion).

The yeast genome contains hundreds of duplications, allowing a comparison of the GC3s of recombinationally hot loci with that of their recombinationally cool paralogs. This approach minimizes the effects of amino acid composition, selective constraints, and gene length on the observed GC3s and reveals that recombinationally hot loci have a significantly greater GC3s

Table 2
Analysis of Mitotic Mismatch Repair Bias

Reference	Mismatch	Total Number Corrected	Number Corrected to GC	Number Corrected to AT	Chi-square values[b]	*P*-value	Correction Bias GC:AT	Base on Nicked Strand
1....	G/T	2,756	2,376	380	1,444	0	6.25	G
1....	G/T	3,321	2,148	1,173	286	0	1.83	G
1....	G/T	1,272	805	467	89	0	1.72	T
1....	G/T	3,150	1,890	1,260	126	0	1.50	T
2....	G/T	1,360	907	453	151	0	2.00	G
2....	G/T	3,310	2,128	1,182	270	0	1.80	G
2....	G/T	1,262	789	473	79	0	1.67	T
3....	G/T	825	382	443	4.4	0.040	0.86	G
3....	G/T	1,963	1,050	913	9.4	0.002	1.15	T
3....	G/T	1,018	496	522	0.6	0.430	0.95	T
3....	G/T	792	392	400	0.1	0.810	0.98	T
5....	G/T	792	466	326	24	8×10^{-7}	1.43	T
5....	G/T	741	395	346	3.2	0.080	1.14	T
5....	G/T	1,199	695	504	30	4×10^{-8}	1.38	T
6....	G/T	136	68	68	0	1.000	1.00	NA
1....	C/A	1,845	1,098	747	66	0	1.47	C
1....	C/A	1,432	747	685	2.6	0.110	1.09	C
1....	C/A	1,307	773	534	43	0	1.45	C
1....	C/A	2,425	1,497	928	133	0	1.61	A
1....	C/A	2,214	1,314	900	77	0	1.46	A
1....	C/A	2,435	1,503	932	133	0	1.61	A
2....	C/A	1,835	1,079	756	57	0	1.43	C
2....	C/A	1,296	648	648	0.0	1.000	1.00	A
2....	C/A	1,398	732	666	21	4×10^{-6}	1.10	A
3....	C/A	1,235	714	521	30	5×10^{-8}	1.37	C
3....	C/A	1,256	754	502	50	0	1.50	C
3....	C/A	835	435	400	1.4	0.240	1.09	C
3....	C/A	425	206	219	0.3	0.560	0.94	A
5....	C/A	925	505	420	7.6	0.006	1.20	C
5....	C/A	746	414	332	8.8	0.003	1.28	C
5....	C/A	1,545	893	652	37	0	1.37	C
6....	C/A	330	189	141	6.7	0.010	1.34	NA
4....	G/A	3,110	1,777	1,333	63	0	1.33	G
4....	G/A	5,446	2,476	2,970	45	0	0.83	A
5....	G/A	148	91	57	7.4	0.007	1.60	G
5....	G/A	583	358	225	30	5×10^{-8}	1.59	G
5....	G/A	422	254	168	17	4×10^{-5}	1.51	G
5....	G/A	168	71	97	3.7	0.054	0.73	A
5....	G/A	843	481	362	17	5×10^{-5}	1.33	A
5....	G/A	261	161	100	14	2×10^{-4}	1.61	A
6....	G/A	116	76	40	11	0.001	1.90	NA
4....	C T	5,464	3,087	2,377	92	0	1.30	C
4....	C T	5,140	2,483	2,657	5.8	0.016	0.94	T
5....	C T	499	338	161	62	0	2.10	C
5....	C T	1,342	868	474	115	0	1.83	C
5....	C T	645	437	208	81	0	2.10	C
5....	C T	199	115	84	4.5	0.034	1.37	T
5....	C T	704	407	297	17	4×10^{-5}	1.37	T
5....	C T	393	214	179	2.9	0.086	1.20	T
6....	C T	107	60	47	1.4	0.245	1.28	NA

[a] References: 1. Yang et al. 1999; 2. Yang et al. 1996; 3. Kang and Kunz 1992; 4. Kunz, Kang, and Kohalmi 1991; 5. Kramer et al. 1989; 6. Bishop, Andersen, and Kolodner 1989.

[b] Chi-square values are rounded off for values greater than 9.5.

(45.5% ± 1.25%) than their nonhot paralogs (37.6% ± 0.85%), Wilcoxon Z = −4.889 ($P = 1.0 \times 10^{-6}$). There is no significant difference, however, between the mean length of these coding sequences, Z = −0.645 (P = 0.52).

It is important to emphasize that the correlations described earlier in this article are not simply broad ranging relationships seen over hundreds of kilobase-pairs but rather occur at a fine scale. This can be visualized in figure 3, which shows a plot of GC3s versus recombination for chromosomes 1–3. As can be seen, GC3s frequently closely mirrors the relative recombination rates of individual ORFs even over short distances encompassing two to four ORFs.

A significant positive correlation was found between the difference in recombinational activity of the

Table 3
Analysis of Meiotic Heteromismatch Correction Bias Within the *HIS4* Locus of *S. cerevisiae* Shows a Significant GC Bias[a]

References[b]	Diploid Strain	Mismatches Examined	Corrected to GC	Corrected to AT	Chi-square values[c]	Correction Bias[d]	*P*-value
1	PD82[e]	T/C and G/A	179	142	4.0	1.26	0.044
1	PD85[e]	T/G and C/A	208	160	6.0	1.30	0.014
1	PD86[e]	T/G and C/A	116	73	9.3	1.59	0.002
1	PD87[e]	T/C and G/A	106	77	4.3	1.38	0.038
1	PD88[e]	T/C and G/A	41	82	13	0.50	2×10^{-4}
1	JS101[e]	T/G and C/A	65	77	0.9	0.84	NS
1	PD11[e]	T/G and C/A	46	29	3.4	1.59	NS
2	PD85[f]	T/G and C/A	111	84	3.5	1.32	NS
2	MW111[f]	T/G and C/A	116	54	22	2.15	3×10^{-6}
3	MW112[f]	T/G and C/A	40	31	0.9	1.29	NS
4	DTK289[f]	T/G and C/A	47	41	0.3	1.15	NS
4	MD50[f]	T/G and C/A	36	35	0	1.03	NS
5	MW104[g]	T/G and C/A	27	26	0	1.04	NS
6	AS4/PD77[e]	T/C and G/A	23	26	0	0.88	NS
6	AS4/PD76[e]	T/G and C/A	25	25	0	1.0	NS
					Mean =	1.22	
			$\Sigma = 1,186$	$\Sigma = 962$			

NOTE.—NS: not significant.

[a] Wilcoxon signed rank test, $Z = -2.04$, ($P = 0.041$).

[b] References: 1. Detloff, Sieber, and Petes 1991; 2. White et al. 1992; 3. Detloff, White, and Petes 1992; 4. Kirkpatrick, Dominska, and Petes 1998; 5. White and Petes 1994; 6. Alani, Reenan, and Kolodner 1994.

[c] Chi-square values greater than 9.5 are rounded off.

[d] Correction bias refers to the ratio of the number of mismatches corrected to GC divided by the number of mismatches corrected to AT.

[e] These experiments examined 6:2, 2:6, 7:1, 1:7 segregation ratios (one conversion event) and 8:0, 0:8 segregation ratios (two conversion events).

[f] These experiments only examined 6:2 and 2:6 segregation ratios.

[g] This experiment examined 6:2 and 2:6 segregation ratios and 8:0 and 0:8 segregation ratios.

hot loci and their nonhot paralogs versus the difference between the hot GC3s and the GC3s of their nonhot paralogs (fig. 4). The difference between the GC3s of the recombinationally active loci and their nonactive paralogs increases significantly as the GC3s of the recombinationally active locus increases (fig. 5). This relationship may reflect different lengths of time since duplication and different lengths of time spent in recombinationally hot and nonhot regions of the genome.

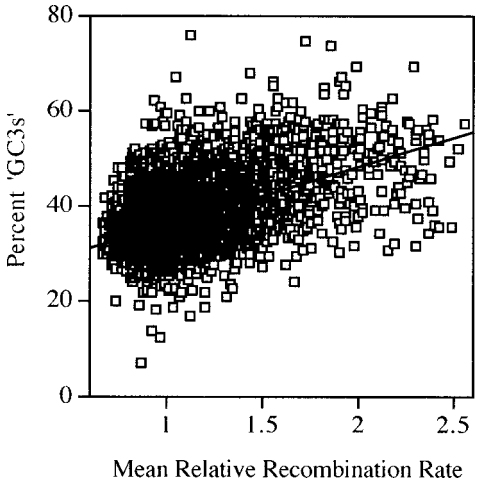

FIG. 1.—Highly significant correlation between GC3s and the average relative rate of recombination of 6,143 open reading frames within the yeast genome $\rho^2 = 0.157$, $Z = 31.0$ ($P = 10^{-211}$). The mean rate of recombination is measured as a function of the array data of Gerton et al. (2000). The 10 most recombinationally active ORFs were omitted from the figure, but not the correlation, for clarity. The line ($y = 12.26x + 23.89$) was fit to the remaining 6,133 open reading frames.

This second set of results demonstrate a highly significant, linear relationship between silent GC content and recombination in *S. cerevisiae*. There are four possible explanations for these observations (1) GC content, per se, may stimulate recombination, (2) selection, (3) mutational bias, and (4) BGC. Each of these hypotheses

Table 4
Spearman Correlation Coefficients Between Various Measures of GC Content and Recombination in 6,143 *S. cerevisiae* Open Reading Frames

Spearman Correlation Between Relative Recombination Rate[a] and:	ρ^2	Z	*P*-value
Total GC.....	0.1109	26.11	1.4×10^{-150}
GC1.....	0.0119	8.57	5.2×10^{-18}
GC2.....	0.0154	9.72	1.2×10^{-22}
GC3.....	0.1560	30.99	3.7×10^{-211}
GC3s.....	0.1568	31.03	1.1×10^{-211}
(GC1 + GC2)/2....	0.0253	12.47	5.4×10^{-36}

[a] Mean recombination rate (as determined from the array data of Gerton et al. 2000).

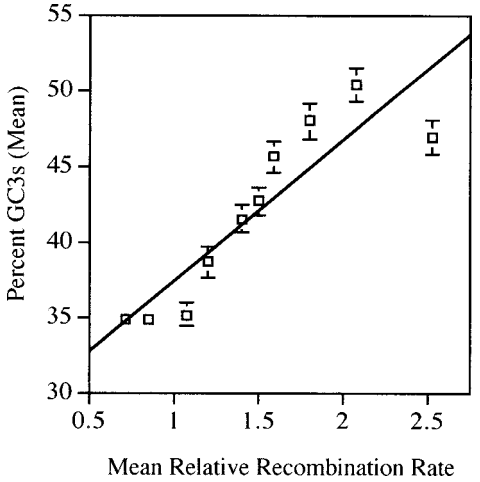

FIG. 2.—Highly significant positive correlation between the GC3s of 500 loci and their relative recombination rate, $\rho^2 = 0.396$, $Z = 14.0$ ($P = 6.8 \times 10^{-45}$). The data point to the far right represents the 50 hottest loci in the genome, and the data point to the far left represents the 50 coldest loci. The equation for the best-fit line is $y = 7.96x + 30.21$.

will be discussed in turn, and evidence will be presented in favor of the fourth hypothesis.

GC Content may Stimulate Recombination

In yeast, most recombination initiating double-strand breaks occur intergenically (Baudat and Nicholas 1997; Gerton et al. 2000), and it has been suggested that the location of these recombination hot spots may be determined, in part, by the GC richness of the adjacent ORFs (Gerton et al. 2000; Petes 2001). If the GC content of the ORFs drives recombination, then the total GC content of the ORFs should explain far more of the variation in recombination rates than GC3s alone. Contrary to this expectation, a Spearman correlation shows that GC3s explains 41% more of the variation in recombination rates and is 61 orders of magnitude more significant than the correlation between total GC content and recombination (table 4). This result is not compatible with a model in which the GC content of ORFs determines their recombination rates but is compatible with BGC.

Galtier et al. (2001) pointed out further evidence, derived from a study by Perry and Ashworth (1999), that recombination drives GC content and not the converse. In mammals, the pseudoautosomal region recombines at a high rate (Ellis and Goodfellow 1989; Blaschke and Rappold 1997). In humans, *R. norvegicus*, and *M. spretus*, the *Fxy* gene is located exclusively on the X chromosome (Perry and Ashworth 1999). In *M. musculos*, the *Fxy* has been rearranged sometime within the past 3 Myr such that the 3′ 1,248 nucleotides are now located in the pseudoautosomal region, leaving 756 nucleotides (GC3s = 63%) on the nonpseudoautosomal X (Ferris et al. 1983; Palmer et al. 1997). This rearrangement was followed by a dramatic increase in the GC content of the pseudoautosomally located *Fxy* (GC3s = 72%) (Perry and Ashworth 1999). The 5′ region of this gene in both *M. musculus* and *M. spretus* is equally divergent from both the rat and human genes. However, in *M. musculus*, the recombinationally hot 3′ end of this gene has experienced a 170-fold greater synonymous substitution rate than the homologous region of the same gene in *M. spretus* (Perry and Ashworth 1999).

To further demonstrate that recombination drives GC content, not the converse, I analyzed the direction of substitutions within the *M. musculus* and *M. spretus Fxy* genes. There were a total of 133 substitutions within *M. musculus Fxy*, and of these, 106 are silent third position changes. Of the 133 substitutions, 127 involved AT to GC substitutions, whereas only 2 involved GC to AT substitutions. In contrast, the *M. spretus* gene has a very low rate of substitution, with only one AT to GC and two GC to AT substitutions. A frequency distribution of the position of these substitutions (fig. 6) shows that the frequency of substitutions increases dramatically at the pseudoautosomal boundary. Within the pseudoautosomal region there were 105 silent third position changes. Of these, 102 were AT to GC, whereas none were in the opposite direction. On the basis of the ancestral GC3 content of the corresponding region (50.1%), the expected number of silent substitutions is 51 AT to GC and 51 GC to AT. The observed number is significantly different, $\chi^2 = 100.0$ ($P = 1.5 \times 10^{-23}$). These observations provide a vivid example of how an increase in the rate of recombination can dramatically increase silent GC content over an evolutionarily brief time span.

Selection

It is very difficult to envision how selection could be responsible for the mouse *Fxy* data because it would imply an enormous mutational load (Galtier et al. 2001). With respect to yeast, if selection is responsible for the positive correlation between recombination and GC3s content, then it should be possible to demonstrate that recombination enhances selection at silent sites, and there should be a significant positive correlation between recombination and codon adaptation. The analysis of 6,143 ORFs shows no significant relationship between GC3s and the codon adaptation index (CAI), $\rho = -0.022$, $Z = -1.73$ ($P = 0.42$), or between recombination and CAI. A scatter plot reveals a distinctive L-shaped distribution, with genes having the highest CAI values being confined primarily to regions of lower recombination (fig. 7). The 500 loci with the highest CAIs actually had lower mean array values than the 500 loci with the lowest CAIs (1.17 ± 0.014 vs. 1.22 ± 0.017), though this difference was not significant. These findings argue strongly against any of the selectionist hypotheses as an explanation for the correlation between GC3s content and recombination. The observation of a positive correlation between CAI and mRNA levels could be caused by the enhanced efficacy of selection brought about by a very large effective population size.

Mutational Bias

It has been suggested that the variation in the GC content could be explained by regional differences in

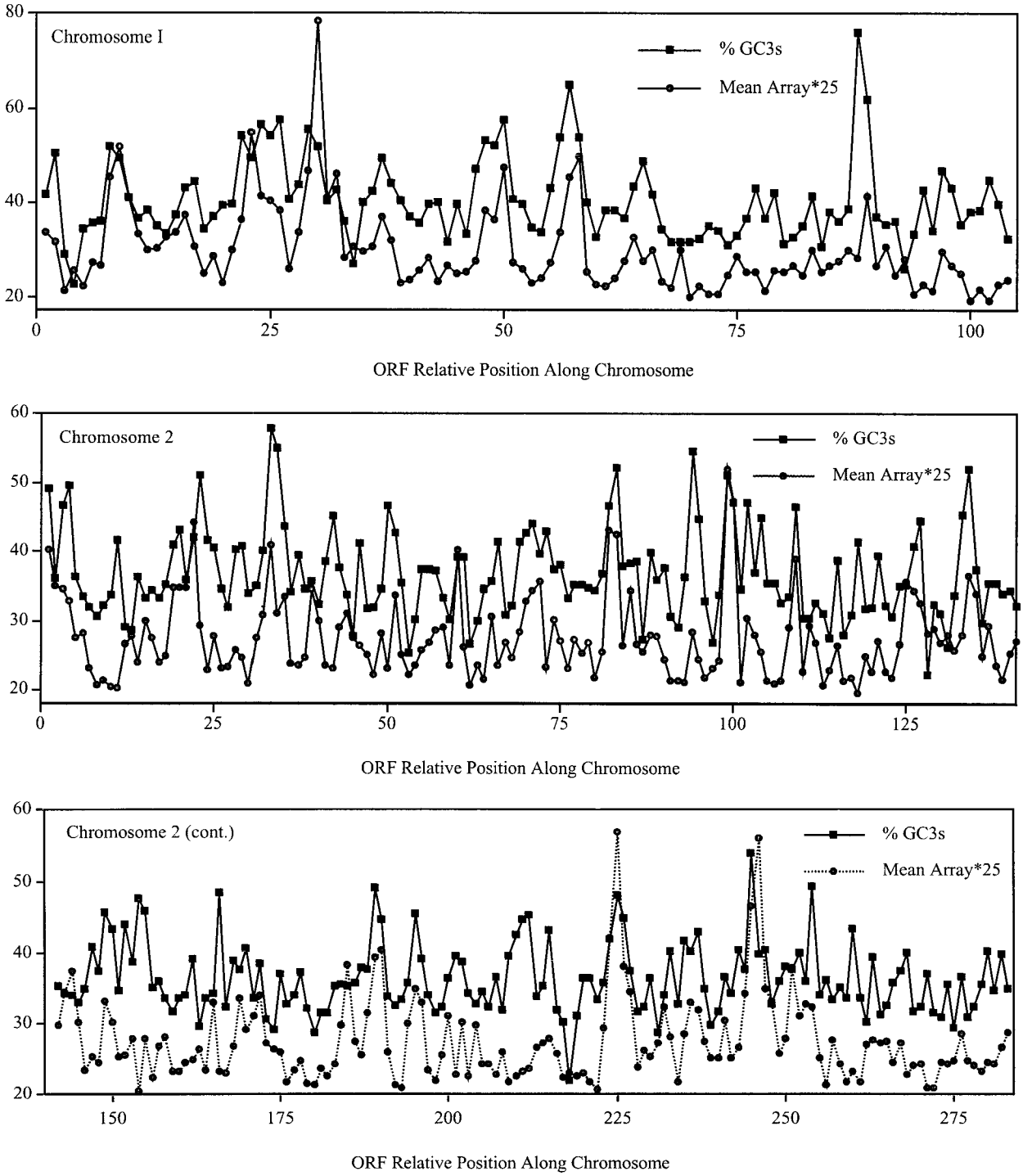

FIG. 3.—Relationship between GC3s and relative recombination rates on *S. cerevisiae* chromosomes 1, 2, and 3. Note that changes in the GC3s often closely mirror changes in the recombination rate over as few as two to four open reading frames. Array values were multiplied by a factor of 25 to aid in graphical presentation.

mutational biases (Sueoka 1962; Filipski 1988; Wolfe, Sharp, and Li 1989; Gu and Li 1994; Francino and Ochman 1999); however, a subsequent analysis has shown that these studies are either not supported by the data or are inconclusive (Eyre-Walker 1992, 1994, 1997, 1999; Eyre-Walker and Hurst 2001; Smith and Eyre-Walker 2001). One significant problem with these mutational hypotheses is that none of them explain the significant relationship between recombination and GC content. Perry and Ashworth (1999) and Marais, Mouchiroud, and Duret (2001) suggested that this correlation might be caused by recombinationally induced GC-biased mutation. Both groups based their conclusions on the results of Strathern, Shafer, and McGill (1995) who dem-

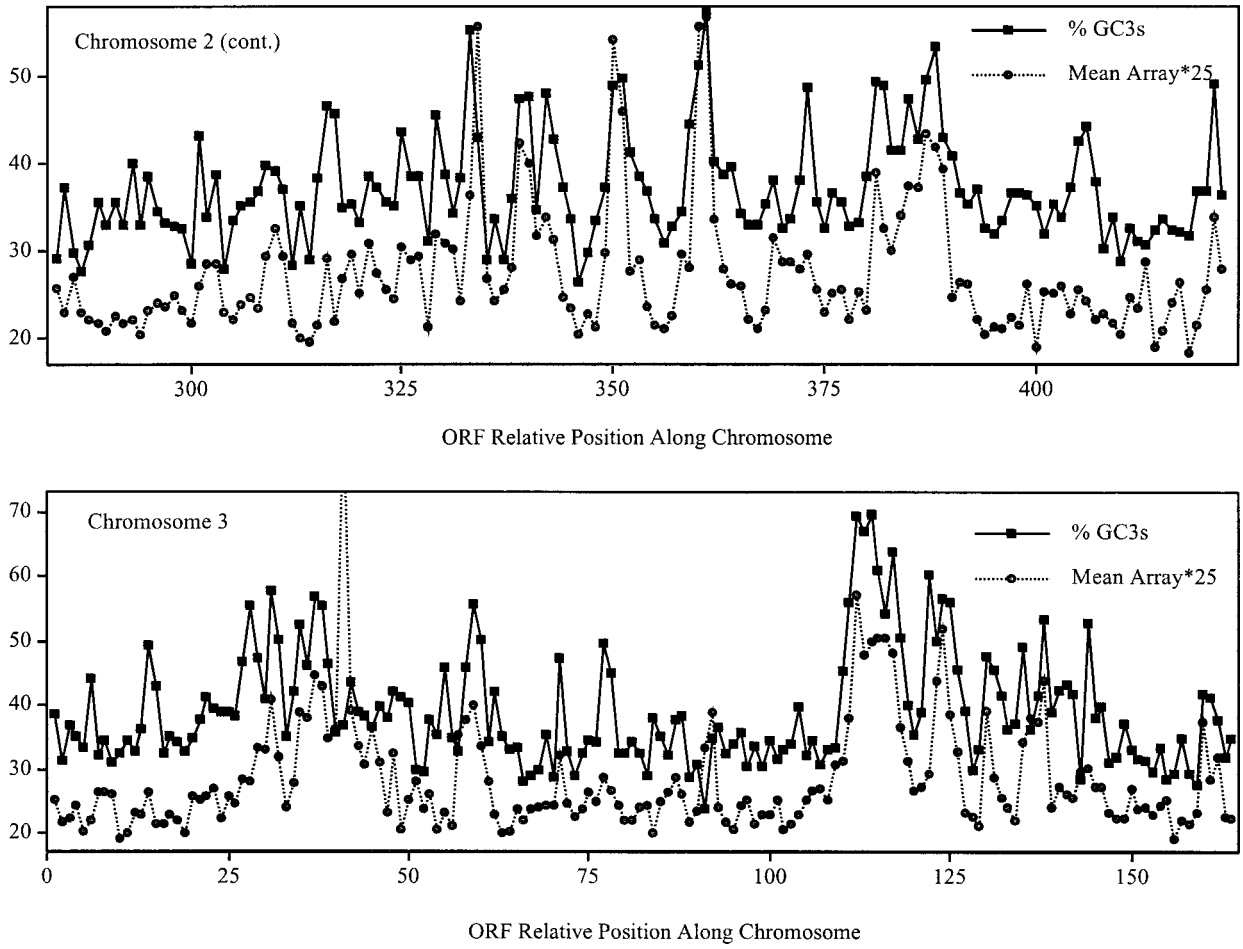

FIG. 3. *(Continued)*

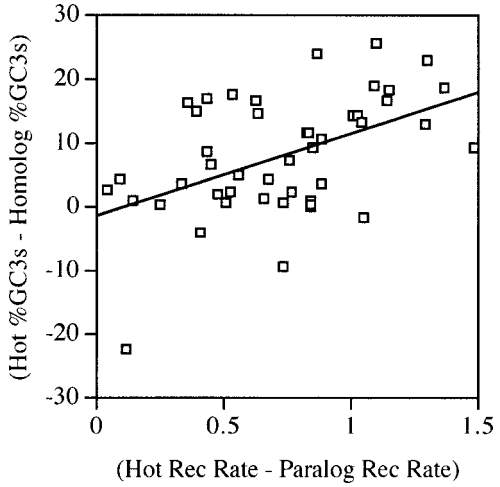

FIG. 4.—Correlation between the difference in the percent GC3s of recombinationally hot and nonhot paralogs and the difference in their relative rates of recombination, $\rho^2 = 0.183$, Z = 2.9 ($P = 0.0019$). One outlier (>5 SD [standard deviation] from the mean) was omitted from the graph, but not the correlation. The line (y = 12.90x–1.40) was fit to the remaining points.

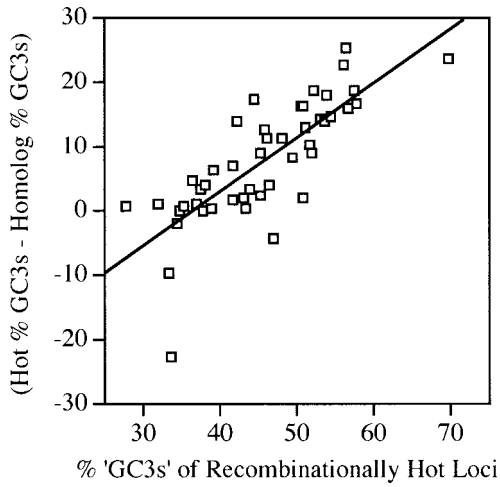

FIG. 5.—Correlation between the difference in the GC3s of the recombinationally hot loci and their recombinationally cool paralogs and GC3s content of the hot loci, $\rho^2 = 0.663$, Z = 5.52 ($P = 1.7 \times 10^{-8}$). One outlier (>5 SD from the mean) was omitted from the graph but not the correlation. The line (y = 0.85–30.71) was fit to the remaining points.

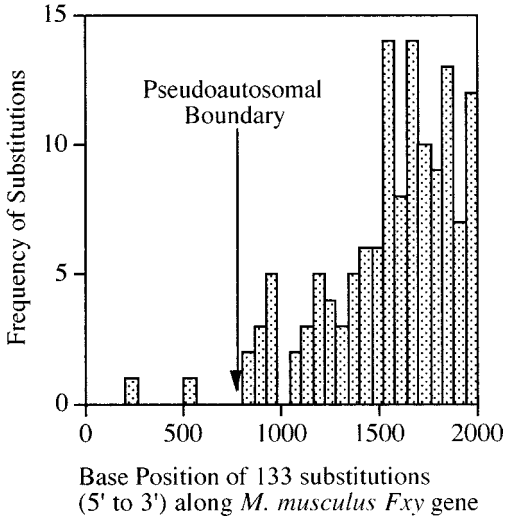

FIG. 6.—Frequency distribution of 133 substitutions within the *M. musculus Fxy* gene. The abscissa represents the position of the substitutions (5′ to 3′) along the gene. Note the position of the pseudoautosomal boundary; substitutions 3′ of this boundary are within the highly recombinagenic pseudoautosomal region.

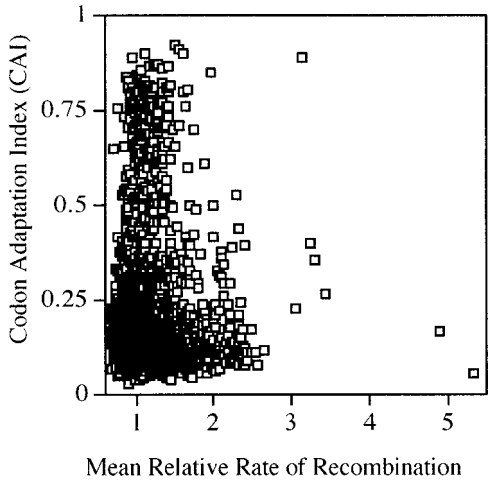

FIG. 7.—Plot of codon adaptation index (CAI) versus the relative rate of recombination for 6,143 yeast open reading frames. Note that there is no significant relationship, $\rho^2 = -0.001$, $Z = -0.04$, ($P = 0.48$).

onstrated that mitotic recombination is mutagenic in yeast. However, an analysis of the results of Strathern, Shafer, and McGill (1995) reveals that of a total of 20 independent, recombination-induced mutations, 12 involved a change from GC to AT, whereas only two involved a change in the reverse direction. Because two different types of AT to GC mutation could be detected using this system, whereas only one type of GC to AT mutation could be detected, the associated χ^2 is 15.0 with a *P*-value of 0.0001 (J. A. Birdsell, unpublished data). Although the data are limited, they show a significant AT mutational bias. This is an important finding, especially if it turns out to be a general phenomenon. Such a bias would provide an even greater selective advantage to the evolution of the GC-biased mismatch repair systems. It must be pointed out that if recombination does have an AT mutational bias, this would in no way contradict the BGC model because the mutation rate is far lower than the rate of biased conversion.

I suggest that another, potentially serious, drawback to the suggestion that recombination causes a GC mutational bias is that in the presence of GC-biased mismatch repair, such a combination would have the potential to greatly increase the mutational load (Bengtsson 1990).

One final problem with the mutational bias hypotheses resides in the fact that the GC content of introns and intergenic regions is typically significantly lower than the GC3s of the genes in which they reside (Aota and Ikemura 1986; D'Onofrio et al. 1991; Clay et al. 1996; Hughes and Yeager 1997; Musto et al. 1999). This is incompatible with a mutational model (Hughes and Yeager 1997; Eyre-Walker 1999). The same holds true for yeast introns, which have a significantly lower GC content (33.8% ± 0.003%) than the GC3s of the corresponding ORFs (38.8% ± 0.005%), $t = 10.4$, df = 220 ($P = 7.3 \times 10^{-21}$) (unpublished data). The mean size of these 228 introns (belonging to 221 ORFs) is 284 ± 14 bp. It is very unlikely that these very short introns could be subject to a different mutational pressure than the exons on either side of them. Identical arguments apply to the intergenic regions of yeast (averaging < 500 bp), which, for the genome as a whole, have an average GC content of 33.16% ± 0.063% as compared with a genomewide average GC3s of 37.00% ± 0.08%, Mann-Whitney $Z = -31.6$ ($P = 3.7 \times 10^{-219}$).

Biased Gene Conversion

Brown and Jiricny (1989) pointed out that GC-biased mismatch repair could lead to an increase in the GC content of recombinationally active regions of the genome. The biased gene conversion (BGC) model does not attempt to explain the location or the reason for the occurrence of recombination initiation hot spots (i.e., the locations where double-strand breaks occur during meiosis). Rather, it explains the observation in numerous organisms of a significant positive correlation between recombination (i.e., regions of heteroduplex formation) and GC content.

The data presented in this paper is totally consistent with the operation of BGC in yeast. If the BGC model is a general phenomenon, then one should be able to demonstrate GC-biased mismatch repair in other organisms showing a positive relationship between recombination and GC content. On the basis of the evidence presented below, I suggest there is a transkingdom GC bias in mismatch repair systems. This would explain the correlations seen in numerous organisms between recombination and GC content. I have compiled evidence of GC-biased mismatch repair in organisms spanning six kingdoms (*sensu* Woese, Kandler, and Wheelis 1990; table 5) (unpublished data). It may not be surprising that these organisms appear to possess GC-biased mismatch repair because evidence suggests that many of them are sub-

Table 5
Some of the Organisms in Which There is Evidence of GC-Biased Mismatch Repair (J.A. Birdsell, unpublished data)

Organism	Method Used To Infer GC Bias (References)[a] and Sequence Accession[b]	Evidence of Positive Relationship Between Recombination and GC? (References)[c]	Evidence of AT Mutational Bias? (References)[d]
Humans[e]	Experimental (1) and AAC50540	Yes (13)	Yes (19)
Simians[e]	Experimental (2)	?	Yes (20)
Chinese Hamsters[e]	Experimental (3)	?	Yes (21)
Mus musculus	Experimental (4) and AAC31900	Yes (14)	Yes (22)
Rattus rattus	AAK08978	Yes (14)	Yes (23)
Bos taurus	TC121174	?	?
Sus scrofa	TC32050	?	?
Gallus gallus	Experimental (5) and AAF14308	Yes (birds) (15)	?
Xenopus laevis[e]	Experimental (6) and TC43299	?	Yes (24)
Danio rerio	AI959337 (cDNA)	?	?
D. melanogaster	Experimental[f] (7) and AAD33588	Yes (16)	Yes (25)
Saccharomyces[e]	Experimental (8)	Yes (17)	Yes (26)
Schizosaccharomyces pombe	AAG01021	?	?
Arabidopsis thaliana	CAB40991	?	?
Zea maize	AW360547 (cDNA)	Yes (18)	Yes (27)
E. coli[e]	Experimental (9) and P17802	?	Yes (28)
Deinococcus radiodurans	Experimental (10) and AAL26976	?	?
Methanothermobacter thermautotrophicus	Experimental (11) P29588	?	?
Pyrobaclum aerophilum	Experimental (12) NP_560564	?	?

NOTE.—Organisms spanning six kingdoms (*sensu* Woese, Kandler, and Wheelis 1990) in which there is evidence of GC-biased mismatch repair. Table also indicates whether there is evidence of a positive relationship between recombination and GC content and whether there is evidence of an AT mutational bias. The accession numbers are for sequences homologous to known GC biased mismatch repair enzymes, such as human G/T specific thymine DNA glycosylase (TDG) or the *E. coli* A/G-specific adenine glycosylase Mut Y. (Birdsell, unpublished data).

[a] References pertaining to GC biased mismatch repair: 1. Brown and Jiricny 1989; Neddermann and Jiricny 1993; Wiebauer and Jiricny 1990; 2. Brown and Jiricny 1988; Heywood and Burke 1990; 3. Bill et al. 1998; Miller et al. 1997; 4. Oda et al. 2000; 5. Zhu et al. 2000; 6. Petranovic et al. 2000; 7. Bhui-kaur, Goodman, and Tower 1998; 8. Birdsell, this article; 9. Au et al. 1988; 10. Li and Lu 2001; 11. Horst and Fritz 1996; 12. Yang et al. 2000.

[b] Accession numbers are for GenBank protein sequences unless otherwise specified. Those numbers beginning with TC are tentative consensus sequences from the TIGR gene index database.

[c] References suggesting a positive relationship between recombination and GC content: 13. Ikemura and Wada 1991; Eyre-Walker 1993; Fullerton, Bernardo Carvalho, and Clark 2001; Birdsell, unpublished data; Eisenbarth et al. 2000; 14. Williams and Hurst 2000; 15. Birdsell, unpublished data; Hurst, Brunton, and Smith 1999; 16. Takano-Shimizu 2001; Marais, Mouchiroud, and Duret 2001; 17. Sharp and Lloyd 1993; Gerton et al. 2000; Birdsell this article, 18. Birdsell unpublished data.

[d] References pertaining to an 'AT' mutational bias: 19. Lander et al. 2001; Gojobori, Li and Graur 1982; 20. Casane et al. 1997; 21. de Jong, Grosovsky, and Glickman 1988; 22. Gojobori, Li, and Graur 1982; Li, Wu, and Luo 1984; 23. Li, Wu and Luo 1984; 24. Gojobori, Li, and Graur 1982; 25. Petrov and Hartl 1999; 26. Sharp and Cowe 1991; Birdsell unpublished data; 27. Birdsell, unpublished data; 28. Halliday and Glickman 1991; Schaaper and Dunn 1991.

[e] Significant GC bias has been experimentally documented.

[f] This experimental evidence is very suggestive of a GC bias but is not conclusive.

ject to an AT-biased mutational pressure (J. A. Birdsell, unpublished data; table 5). Such a mutational pressure can result from a variety of fundamental processes, including the spontaneous deamination of cytosine to Uracil or 5-methylcytosine to thymine (Coulondre et al. 1978; Duncan and Miller 1980), oxidative damage to cytosine (Kreutzer and Essigmann 1998) or guanine (Newcomb and Loeb 1998), or UV irradiation (Peng and Shaw 1996), all of which can result in GC to AT or TA mutations. The fact that virtually every organism in which a correlation has been found between recombination and GC content appears to possess a GC-biased mismatch repair system provides strong circumstantial evidence in favor of the BGC model. Recent articles by Eyre-Walker and Hurst (2001) and Galtier et al. (2001) provide additional support for the BGC model.

Potential Drawbacks of the BGC Model

There are several potential problems with the BGC model (Eyre-Walker 1999; Eyre-Walker and Hurst 2001)

Table 6
Human Genes Located on the X Chromosome Have Significantly Greater Silent GC Content than Their Y Homologs[a]

X Homolog	X Homolog GC3s	Y Homolog GC3s	Y Homolog	Million Years w/o Rec.[b]
PKXI	74.42	55.95	*PRKY*	30–50
AMELX	61.93	59.22	*AMELY*	30–50
Thymosin beta-4	50.00	45.24	*Thymosin beta-4 Y*	80–130
ZFX	40.00	35.73	*ZFY*	80–130
eIF-4C	32.14	30.50	*eIF-1A, Y*	80–130
DFFRX	31.48	29.26	*DFFRY*	80–130
DBX	36.97	31.09	*DBY*	80–130
UTX	37.25	32.87	*UTY*	80–130
SMCX	66.76	56.01	*SMCY*	130–170
RPS4X	52.73	51.56	*RPS4Y*	240–320
RMBX	25.33	22.77	*RMBY*	240–320
SOX3	79.43	56.54	*SRY*	240–320
TBL1	61.73	53.63	*TBL1Y*	Unknown
VCX	69.80	70.73	*VCY*	Unknown
PCDHX	40.21	39.16	*PCDHY*	Unknown
Mean ± S.E.	50.68 ± 4.48	44.68 ± 3.63		

[a] The GC3s of 15 pairs of human X-Y homologs, along with the estimated number of years that the Y homolog has been without recombination according to Lahn and Page (1999). No pseudogenes were included in this analysis. A paired *t*-test yielded a *t* of 3.47 ($P = 0.0038$). There is no relationship between the difference in GC3s content of the two homologs and the length of time without recombination.

[b] Data from Lahn and Page (1999).

(1) K_s (the synonymous substitution rate) may (depending upon the method used to calculate it) positively covary with the GC4 content (where GC4 is the GC content at fourfold degenerate sites) (Hurst and Williams 2000), (2) the model is highly parameter sensitive, (3) there are ancient Y-linked loci such as *SRY* that have relatively high GC contents, and (4) introns typically have lower GC content than neighboring exons. With respect to the first potential drawback, I suggest that if the BGC model is correct, then algorithms which fail to incorporate BGC into their calculations of K_a and K_s may lead to inaccurate estimates of these parameters. As for the second potential problem, it has been stated that there is only ''a one order of magnitude window'' within which BGC can function (Eyre-Walker 1999). Galtier et al. (2001) argue, however, that this does not pose a serious problem because in real populations extremely high levels of BGC would probably be selected against. With respect to the third problem, although some Y-linked loci are indeed fairly GC rich, I suggest they would have an even higher GC content if they were autosomally located or located on the X chromosome. Data pertaining to this are presented in table 6. Whereas *SRY* does have a GC3s of 56.5%, its X homolog has a GC3s of 79.4%. Overall, the human X homologs have a significantly higher GC3s (50.6 ± 4.5) than the Y homologs (44.7 ± 3.6), $t = 3.47$ ($P = 0.0038$), and as can be seen, in every instance, except for one, the X homolog has a higher GC3s than its Y counterpart.

The observation that introns have a lower GC content than the neighboring exons appears difficult to reconcile with the BGC model (Eyre-Walker 1999). There are, however, at least two models that could explain this observation. According to one model, introns have lower GC content because they are the preferred sites for the insertion of transposable elements, which, in some organisms, typically have lower GC content than the regions into which they insert (Duret and Hurst 2001). Although this is an ingenious hypothesis to explain the lower GC content of vertebrate introns, it cannot explain the lower GC content of yeast introns or intergenic regions which, having average lengths of 284 and 484 bp, respectively, are far too small to house even one Ty element. The second model is presented below.

The Constraint Model

The intergenic regions of many organisms have, on average, lower GC content than the silent GC content of ORFs on either side of them (Clay et al. 1996). Pseudogenes also usually have a lower GC content than their functional counterparts (Gojobori, Li, and Graur 1982; Li, Wu, and Luo 1984; Petrov and Hartl 1999). Here I propose a model, referred to as the Constraint hypothesis, which may explain the lower GC content of introns, intergenic regions, and pseudogenes. This model is based upon the observation that the greater the heterology between two sequences, the less likely it is that recombination will occur between them. This ''antirecombinagenic'' effect of sequence heterology has been well documented in prokaryotes (Shen and Huang 1989; Roberts and Cohan 1993; Vulic et al. 1997), yeast (Borts and Haber 1987; Datta et al. 1997; Chen and Jinks-Robertson 1999), and mammalian cells (Waldman and Liskay 1988; Lukacsovich and Waldman 1999), and mismatch repair enzymes have been shown to be responsible for preventing recombination between diverged sequences in a variety of organisms (Rayssiguier, Thaler, and Radman 1989; Borts et al. 1990; Chen and Jinks-Robertson 1999).

Nonregulatory, noncoding regions of the genome are under less selective constraint than regulatory or

coding regions; therefore, they evolve more rapidly and have higher levels of polymorphism (Hughes and Yeager 1997; Shabalina et al. 2001). I suggest that, on average, heteroduplex formation and propagation should be expected to occur most frequently within conserved coding and regulatory regions of the genome. At the population level, these more conserved regions of the genome will possess lower levels of polymorphism, which, in an outcrossing organism, translates into less heterology within the individual. This could explain why the GC3s of coding regions and the GC content of regulatory regions (Babenko et al. 1999; unpublished data) is higher than the GC content of introns, intergenic regions, or pseudogenes. The analysis of sequences from 697 yeast regulatory elements (belonging to 99 different element types) shows that they have a significantly higher mean GC content (45.72 $\pm$ 0.66) than yeast intergenic regions as a whole (33.16 $\pm$ 0.06) (Mann-Whitney $Z = -23.3$; $P = 4.4 \times 10^{-120}$). The mean GC content of these 99 types of regulatory element (47.47 $\pm$ 1.53) is also significantly greater than that of intergenic regions, $Z = -11.9$ ($P = 1.2 \times 10^{-32}$). These 697 regulatory sequences also have a significantly higher GC content than the mean GC3s of 6,330 ORFs (37.00 $\pm$ 0.08), $Z = -16.81$ ($P = 2.0 \times 10^{-63}$), as do the 99 types of element, $Z = -9.1$ ($P = 9.0 \times 10^{-20}$).

The process proposed by the Constraint model would lead to a positive feedback loop in which selective constraint leads to increased rates of recombination, which in turn would enhance the efficacy of selection, thereby increasing the selective constraint. The Constraint hypothesis does not seek to explain the cause or location of recombination initiation hot spots. Rather, it seeks to point out that, given there is a recombination hot spot, heteroduplex formation and propagation will, on average, proceed from this hot spot into the conserved coding or regulatory regions more frequently than into nonconserved regions. An important implication of the Constraint hypothesis is that the large intergenic and intronic regions of organisms such as humans would not contribute proportionately to the genetic map size of such organisms. Further support for this Constraint model comes from a number of independent sources and will be presented in detail elsewhere.

Conclusions

A number of lines of evidence are presented in this paper in support of the BGC model. There is a highly significant positive correlation between recombination and silent GC content in the yeast *S. cerevisiae*. This relationship cannot be explained by any of the other models examined. Any model attempting to explain regional variations in GC content must not only explain the relationship between GC content and recombination but also the observation that GC3s is almost always higher than the GC content of introns, pseudogenes, or intergenic regions. The BGC model, in conjunction with the Constraint model, can do so. For the first time in any member of the fungi kingdom, a significant GC-biased mismatch repair system is found operating in both mitotic as well as meiotic cells. This repair bias may have evolved in response to the AT mutational bias to which *S. cerevisiae* is subjected. Much of the variation in the GC content within the yeast genome may therefore be a result of the interplay between AT-biased mutational pressure and GC-biased gene conversion.

Evidence suggests that a number of other organisms spanning several kingdoms may be subjected to similar processes. Virtually all organisms in which a correlation exists between recombination and GC content also appear to possess GC-biased mismatch repair. I suggest that this transkingdom GC bias in mismatch repair systems has evolved in response to a prevailing AT mutational bias resulting from fundamental properties of DNA.

Nonrecombining regions of the genome and nonrecombining genomes would not be subject to the molecular drive caused by BGC. I suggest that the low GC content, characteristic of nonrecombining genomes, may be the result of three processes: (1) a prevailing AT mutational bias, (2) random fixation of the most common types of mutation caused by genetic drift, and (3) the absence of GC-biased gene conversion which, in recombining organisms, would permit the reversal of the most common form of mutation.

A model is presented to explain the observations that the GC content of introns, pseudogenes, and intergenic regions is almost always lower than the silent GC content of open reading frames. According to this Constraint model, heteroduplex formation and propagation is expected to occur, on average, more frequently within the regions of the genome that are under greater selective constraint, such as conserved regulatory and coding regions. The higher GC content of such conserved regions supports this view. In summary, I suggest that much of the variation in GC content seen in organisms spanning several kingdoms may be attributed, in part, to the interplay between a prevailing AT mutational pressure, recombination, GC-biased mismatch repair, and the antirecombinagenic effects of sequence heterology.

Because most point mutations are GC to AT events, recombination allows the most common form of mutation to be restored to wild type through the actions of GC-biased mismatch repair. In recombining organisms, mismatches occur through mutation as well as through recombination. In nonrecombining organisms, mismatches only occur through mutation. Recombination therefore provides mismatch repair enzymes multiple chances to repair the most common type of mutation. Nonrecombining organisms would not be afforded such opportunities. I suggest that this ability to resurrect wild-type alleles from mutant alleles would have powerful and immediate selective advantages through its potential to reduce both the number of mutations within the recombining genome as well the mutational load of the outcrossing population. This Mutation Reversal model may explain, in part, the evolution of several forms of sexual recombination, including meiotic recombination and genetic transformation, and will be presented in detail elsewhere. For those interested in a comprehensive

review of other contemporary models for the evolution of sex see Birdsell and Wills 2001.

The findings presented here have implications for a variety of fields of research. With respect to DNA repair, it appears that *S. cerevisiae* may posses both a thymine and an adenine DNA glycosylase activity. No such enzymes have ever been characterized in this organism, and blast-p searches of known thymine and adenine glycosylases against the *S. cerevisiae* genome have turned up no candidate loci, suggesting that genes of uncharacterized function are responsible for these mismatch repair activities.

Given the paucity of accurate data on the recombination rates in organisms such as humans, silent GC content may be a useful first order approximation of the relative recombination rate of a locus. Algorithms used to calculate evolutionary parameters, such as K_a and K_s, may benefit by taking into account the recombinational background of loci as well as the effect and degree of BGC on estimates of K_a and K_s. I suggest that the theory of directional mutation pressure (Sueoka 1962) may require modification such that it applies only to selectively neutral regions of the genome which are not subject to BGC.

Phylogenetic models may also benefit by taking into consideration the recombinational background of the loci under investigation. Failure to do so may result in an underestimate of divergence in recombinationally hot loci because of mutation reversal as well as convergent evolution (i.e., if GC mutations have a greater chance of fixation, then two sequences may appear more similar because of a common form of molecular drive acting on them). Models of the evolution of sex may benefit by incorporating the possibility that recombination helps reverse the most common form of mutation through GC-biased gene conversion.

Acknowledgments

I would like to thank the following people for their assistance and helpful comments: Margaret Kidwell, Ken Wolfe, Eric Alani, Bruce Walsh, Rick Michod, Bill Birky, Bernard Kunz, Chris Wills, Dawn Birdsell, Megan McCarthy, Tassia Kolesnikow, Lillian Engel, Ted Weinert, Tom Petes, and two anonymous reviewers. I would also like to thank James McInerney, Ziheng Yang, and Etsuko Moriyama for kindly making their software available.

LITERATURE CITED

ALANI, E., R. A. REENAN, and R. D. KOLODNER. 1994. Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. Genetics **137**:19–39.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG, W. MILLER, and D. J. LIPMAN. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

AOTA, S., and T. IKEMURA. 1986. Diversity in G+C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res. **14**:6345–6355.

AU, K. G., M. CABRERA, J. H. MILLER, and P. MODRICH. 1988. *Escherichia coli* mutY gene product is required for specific A-G–C. G mismatch correction. Proc. Natl. Acad. Sci. USA **85**:9163–9166.

BABENKO, V. N., P. S. KOSAREV, O. V. VISHNEVSKY, V. G. LEVITSKY, V. V. BASIN, and A. S. FROLOV. 1999. Investigating extended regulatory regions of genomic DNA sequences. Bioinformatics **15**:644–653.

BAUDAT, F., and A. NICHOLAS. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. Proc. Natl. Acad. Sci. USA **94**:5213–5218.

BENGTSSON, B. O. 1990. The effect of biased conversion on the mutation load. Genet. Res. **55**:183–187.

BERNARDI, G. 1986. Compositional constraints and genome evolution. J. Mol. Evol. **24**:1–11.

———. 2000. Isochores and the evolutionary genomics of vertebrates. Gene **241**:3–17.

BHUI-KAUR, A., M. F. GOODMAN, and J. TOWER. 1998. DNA mismatch repair catalyzed by extracts of mitotic, postmitotic, and senescent *Drosophila* tissues and involvement of *mei-9* gene function for full activity. Mol. Cell. Biol. **18**: 1436–1443.

BILL, C. A., W. A. DURAN, N. R. MISELIS, and J. A. NICKOLOFF. 1998. Efficient repair of all types of single-base mismatches in recombination intermediates in Chinese hamster ovary cells: competition between long-patch and G-T glycosylase–mediated repair of G-T mismatches. Genetics **149**: 1935–1943.

BIRDSELL, J. A., and C. WILLS. 2001. The evolutionary origin and maintenance of sexual recombination: a review of contemporary models. Evol. Biol. (in press).

BISHOP, D. K., J. ANDERSEN, and R. D. KOLODNER. 1989. Specificity of mismatch repair following transformation of *Saccharomyces cerevisiae* with heteroduplex plasmid DNA. Proc. Natl. Acad. Sci. USA **86**:3713–3717.

BLASCHKE, R. J., and G. A. RAPPOLD. 1997. Man to mouse—lessons learned from the distal end of the human X chromosome. Genome Res. **7**:1114–1117.

BORTS, R. H., and J. E. HABER. 1987. Meiotic recombination in yeast: alteration by multiple heterzygosities. Science **237**: 1459–1465.

BORTS, R. H., W. Y. LEUNG, W. KRAMER, B. KRAMER, M. WILLIAMSON, S. FOGEL, and J. E. HABER. 1990. Mismatch repair–induced meiotic recombination requires the pms1 gene product. Genetics **124**:573–584.

BRADNAM, K. R., C. SEOIGHE, P. M. SHARP, and K. H. WOLFE. 1999. G+C content variation along and among *Saccharomyces cerevisiae* chromosomes. Mol. Biol. Evol. **16**:666–675.

BROWN, T. C., and J. JIRICNY. 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell **54**:705–711.

———. 1989. Repair of base-base mismatches in simian and human cells. Genome **31**:578–583.

CASANE, D., S. BOISSINOT, B. H. CHANG, L. C. SHIMMIN, and W. LI. 1997. Mutation pattern variation among regions of the primate genome. J. Mol. Evol. **45**:216–226.

CHARLESWORTH, B. 1994. Patterns in the genome. Curr. Biol. **4**:182–184.

CHEN, W., and S. JINKS-ROBERTSON. 1999. The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. Genetics **151**:1299–1313.

CLAY, O., S. CACCIO, S. ZOUBAK, D. MOUCHIROUD, and G. BERNARDI. 1996. Human coding and noncoding DNA: compositional correlations. Mol. Phylogenet. Evol. **5**:2–12.

COULONDRE, C., J. H. MILLER, P. J. FARABAUGH, and W. GILBERT. 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. Nature **274**:775–780.

DATTA, A., M. HENDRIX, M. LIPSITCH, and S. JINKS-ROBERTSON. 1997. Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. Proc. Natl. Acad. Sci. USA **94**:9757–9762.

D'ONOFRIO, G., D. MOUCHIROUD, B. AISSANI, C. GAUTIER, and G. BERNARDI. 1991. Correlations between the compositional properties of human genes, codon usage, and amino acid composition of proteins. J. Mol. Evol. **32**:504–510.

DE JONG, P. J., A. J. GROSOVSKY, and B. W. GLICKMAN. 1988. Spectrum of spontaneous mutation at the *APRT* locus of Chinese hamster ovary cells: an analysis at the DNA sequence level. Proc. Natl. Acad. Sci. USA **85**:3499–3503.

DETLOFF, P., J. SIEBER, and T. D. PETES. 1991. Repair of specific base pair mismatches formed during meiotic recombination in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **11**:737–745.

DETLOFF, P., M. A. WHITE, and T. D. PETES. 1992. Analysis of a gene conversion gradient at the *HIS4* locus in *Saccharomyces cerevisiae*. Genetics **132**:113–123.

DUNCAN, B. K., and J. H. MILLER. 1980. Mutagenic deamination of cytosine residues in DNA. Nature **287**:560–561.

DURET, L., and L. D. HURST. 2001. The elevated GC content at exonic third sites is not evidence against neutralist models of isochore evolution. Mol. Biol. Evol. **18**:757–762.

EISENBARTH, I., A. M. STRIEBEL, E. MOSCHGATH, W. VOGEL, and G. ASSUM. 2001. Long-range sequence composition mirrors linkage disequilibrium pattern in a 1.13 Mb region of human chromosome 22. Hum. Mol. Genet. **10**:2833–2839.

EISENBARTH, I., G. VOGEL, W. KRONE, W. VOGEL, and G. ASSUM. 2000. An isochore transition in the *NF1* gene region coincides with a switch in the extent of linkage disequilibrium. Am. J. Hum. Genet. **67**:873–880.

ELLIS, N., and P. N. GOODFELLOW. 1989. The mammalian pseudoautosomal region. Trends Genet. **5**:406–410.

EYRE-WALKER, A. 1992. Evidence that both G + C rich and G + C poor isochores are replicated early and late in the cell cycle. Nucleic Acids Res. **20**:1497–1501.

———. 1993. Recombination and mammalian genome evolution. Proc. R. Soc. Lond. B **252**:237–243.

———. 1994. DNA mismatch repair and synonymous codon evolution in mammals. Mol. Biol. Evol. **1**:88–98.

———. 1997. Differentiating between selection and mutation bias. Genetics **147**:1983–1987.

———. 1999. Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics **152**:675–683.

EYRE-WALKER, A., and L. D. HURST. 2001. OPINION: the evolution of isochores. Nat. Rev. Genet. **2**:549–555.

FERRIS, S. D., R. D. SAGE, E. M. PRAGER, U. RITTE, and A. C. WILSON. 1983. Mitochondrial DNA evolution in mice. Genetics **105**:681–721.

FILIPSKI, J. 1987. Correlation between molecular clock ticking, codon usage fidelity of DNA repair, chromosome banding and chromatin compactness in germline cells. FEBS Lett. **217**:184–186.

———. 1988. Why the rate of silent codon substitutions is variable within a vertebrate's genome. J. Theor. Biol. **134**:159–164.

FRANCINO, M. P., and H. OCHMAN. 1999. Isochores result from mutation not selection. Nature **400**:30–31.

FULLERTON, S. M., A. BERNARDO CARVALHO, and A. G. CLARK. 2001. Local rates of recombination are positively correlated with GC content in the human genome. Mol. Biol. Evol. **18**:1139–1142.

GALTIER, N., G. PIGANEAU, D. MOUCHIROUD, and L. DURET. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics **159**:907–911.

GERTON, J. L., J. DERISI, R. SHROFF, M. LICHTEN, P. O. BROWN, and T. D. PETES. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **97**:11383–11390.

GOJOBORI, T., W. H. LI, and D. GRAUR. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. **18**:360–369.

GRILLO, G., M. ATTIMONELLI, S. LIUNI, and G. PESOLE. 1996. CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. Comput. Appl. Biosci. **12**:1–8.

GU, X., and W. H. LI. 1994. A model for the correlation of mutation rate with GC content and the origin of GC-rich isochores. J. Mol. Evol. **38**:468–475.

HALLIDAY, J. A., and B. W. GLICKMAN. 1991. Mechanisms of spontaneous mutation in DNA repair-proficient *Escherichia coli*. Mutat. Res. **250**:55–71.

HEYWOOD, L. A., and J. F. BURKE. 1990. Mismatch repair in mammalian cells. Bioessays **12**:473–477.

HOLMQUIST, G. P. 1992. Chromosome bands, their chromatin flavors, and their functional features. Am. J. Hum. Genet. **51**:17–37.

HORST, J. P., and H. J. FRITZ. 1996. Counteracting the mutagenic effect of hydrolytic deamination of DNA 5-methylcytosine residues at high temperature: DNA mismatch *N*-glycosylase Mig.Mth of the thermophilic archaeon *Methanobacterium thermoautotrophicum* THF. EMBO J. **15**:5459–5469.

HUGHES, A. L., and M. YEAGER. 1997. Comparative evolutionary rates of introns and exons in murine rodents. J. Mol. Evol. **45**:125–130.

HURST, L. D., C. F. BRUNTON, and N. G. SMITH. 1999. Small introns tend to occur in GC-rich regions in some but not all vertebrates. Trends Genet. **15**:437–439.

HURST, L. D., and E. J. WILLIAMS. 2000. Covariation of GC content and the silent site substitution rate in rodents: implications for methodology and for the evolution of isochores. Gene **261**:107–114.

IKEMURA, T., and K. WADA. 1991. Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers; relation between nucleotide sequence data and cytogenetic data. Nucleic Acids Res. **19**:4333–4339.

KANG, X., and B. A. KUNZ. 1992. Inactivation of the *RAD1* excision-repair gene does not affect correction of mismatches on heteroduplex plasmid DNA in yeast. Curr. Genet. **21**:261–263.

KIRKPATRICK, D. T., M. DOMINSKA, and T. D. PETES. 1998. Conversion-type and restoration-type repair of DNA mismatches formed during meiotic recombination in *Saccharomyces cerevisiae*. Genetics **149**:1693–1705.

KRAMER, B., W. KRAMER, M. S. WILLIAMSON, and S. FOGEL. 1989. Heteroduplex DNA correction in *Saccharomyces cerevisiae* is mismatch specific and requires functional *PMS* genes. Mol. Cell. Biol. **9**:4432–4440.

KREUTZER, D. A., and J. M. ESSIGMANN. 1998. Oxidized, deaminated cytosines are a source of C to T transitions in vivo. Proc. Natl. Acad. Sci. USA **95**:3578–3582.

KUNZ, B. A., X. L. KANG, and L. KOHALMI. 1991. The yeast rad18 mutator specifically increases G.C→T.A transversions without reducing correction of G-A or C-T mismatches to G.C pairs. Mol. Cell. Biol. **11**:218–225.

LAHN, B. T., and D. C. PAGE. 1999. Four evolutionary strata on the human X chromosome. Science **286**:964–967.

LANDER, E. S., L. M. LINTON, B. BIRREN et al. (248 co-authors). 2001. Initial sequencing and analysis of the human genome. Nature **409**:860–921.

LI, X., and A. L. LU. 2001. Molecular cloning and functional analysis of the MutY homolog of *Deinococcus radiodurans*. J. Bacteriol. **183**:6151–6158.

LI, W. H., C. I. WU, and C. C. LUO. 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J. Mol. Evol. **21**:58–71.

LUKACSOVICH, T., and A. S. WALDMAN. 1999. Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. Genetics **151**:1559–1568.

MARAIS, G., D. MOUCHIROUD, and L. DURET. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc. Natl. Acad. Sci. USA **98**:5688–5692.

MATASSI, G., L. M. MONTERO, J. SALINAS, and G. BERNARDI. 1989. The isochore organization and the compositional distribution of homologous coding sequences in the nuclear genome of plants. Nucleic Acids Res. **17**:5273–5290.

MCINERNEY, J. O. 1998. GCUA: general codon usage analysis. Bioinformatics **14**:372–373.

MILLER, E. M., H. L. HOUGH, J. W. CHO, and J. A. NICKOLOFF. 1997. Mismatch repair by efficient nick-directed, and less efficient mismatch-specific, mechanisms in homologous recombination intermediates in Chinese hamster ovary cells. Genetics **147**:743–753.

MUSTO, H., H. ROMERO, A. ZAVALA, and G. BERNARDI. 1999. Compositional correlations in the chicken genome. J. Mol. Evol. **49**:325–329.

NEDDERMANN, P., and J. JIRICNY. 1993. The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. J. Biol. Chem. **268**:21218–21224.

NEKRUTENKO, A., and W.-H. LI. 2000. Assessment of compositional heterogeneity within and between eukaryotic geneomes. Genome Res. **10**:1986–1995.

NEWCOMB, T. G., and L. A. LOEB. 1998. Oxidative DNA damage and mutagenesis. Pp. 65–84 *in* J. A. NICKOLOFF and M. F. HOEKSTRA, eds. DNA damage and repair: DNA repair in prokaryotes and lower eukaryotes. Humana, Totowa, NJ.

ODA, S., O. HUMBERT, S. FIUMICINO, M. BIGNAMI, and P. KARRAN. 2000. Efficient repair of A/C mismatches in mouse cells deficient in long-patch repair. EMBO J. **19**:1711–1718.

PALMER, S., J. PERRY, D. KIPLING, and A. ASHWORTH. 1997. A gene spans the pseudoautosomal boundary in mice. Proc. Natl. Acad. Sci. USA **94**:12030–12035.

PEARSON, W. R. 1991. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics **11**:635–650.

PENG, W., and B. R. SHAW. 1996. Accelerated deamination of cytosine residues in UV-induced cyclobutane pyrimidine dimers leads to CC→TT transitions. Biochemistry **35**:10172–10181.

PERRY, J., and A. ASHWORTH. 1999. Evolutionary rate of a gene affected by chromosomal position. Curr. Biol. **9**:987–989.

PETES, T. D. 2001. Meiotic recombination hot spots and cold spots. Nat. Rev. Genet. **2**:360–369.

PETRANOVIC, M., K. VLAHOVIC, D. ZAHRADKA, S. DZIDIC, and M. RADMAN. 2000. Mismatch repair in *Xenopus* egg extracts is not strand-directed by DNA methylation. Neoplasma **47**:375–381.

PETROV, D. A., and D. L. HARTL. 1999. Patterns of nucleotide substitution in *Drosophila* and mammalian genomes. Proc. Natl. Acad. Sci. USA **96**:1475–1479.

PROFFITT, J. H., J. R. DAVIE, D. SWINTON, and S. HATTMAN. 1984. 5-Methylcytosine is not detectable in *Saccharomyces cerevisiae* DNA. Mol. Cell. Biol. **4**:985–988.

RAYSSIGUIER, C., D. S. THALER, and M. RADMAN. 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature **342**:396–401.

ROBERTS, M. S., and F. M. COHAN. 1993. The effect of DNA sequence divergence on sexual isolation in *Bacillus*. Genetics **134**:401–408.

SCHAAPER, R. M., and R. L. DUNN. 1991. Spontaneous mutation in the *Escherichia colilacI* gene. Genetics **129**:317–326.

SHABALINA, S. A., A. Y. OGURTSOV, V. A. KONDRASHOV, and A. S. KONDRASHOV. 2001. Selective constraint in intergenic regions of human and mouse genomes. Trends Genet. **17**:373–376.

SHARP, P. M., and E. COWE. 1991. Synonymous codon usage in *Saccharomyces cerevisiae*. Yeast **7**:657–678.

SHARP, P. M., and A. T. LLOYD. 1993. Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. Nucleic Acids Res. **21**:179–183.

SHEN, P., and H. V. HUANG. 1989. Effect of base pair mismatches on recombination via the RecBCD pathway. Mol. Gen. Genet. **218**:358–360.

SMITH, N. G., and A. EYRE-WALKER. 2001. Synonymous codon bias is not caused by mutation bias in G+C-rich genes in humans. Mol. Biol. Evol. **18**:982–986.

STRATHERN, J. N., B. K. SHAFER, and C. B. MCGILL. 1995. DNA synthesis errors associated with double-strand-break repair. Genetics **140**:965–972.

SUEOKA, N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA **48**:582–592.

TAKANO-SHIMIZU, T. 2001. Local changes in GC/AT substitution biases and in crossover frequencies on *Drosophila* chromosomes. Mol. Biol. Evol. **18**:606–619.

VULIC, M., F. DIONISIO, F. TADDEI, and M. RADMAN. 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc. Natl. Acad. Sci. USA **94**:9763–9767.

WALDMAN, A. S., and R. M. LISKAY. 1988. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol. Cell. Biol. **8**:5350–5357.

WHITE, M. A., and T. D. PETES. 1994. Analysis of meiotic recombination events near a recombination hotspot in the yeast *Saccharomyces cerevisiae*. Curr. Genet. **26**:21–30.

WHITE, M. A., P. DETLOFF, M. STRAND, and T. D. PETES. 1992. A promoter deletion reduces the rate of mitotic, but not meiotic, recombination at the *HIS4* locus in yeast. Curr. Genet. **21**:109–116.

WIEBAUER, K., and J. JIRICNY. 1990. Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. Proc. Natl. Acad. Sci. USA **87**:5842–5845.

WILLIAMS, E. J., and L. D. HURST. 2000. The proteins of linked genes evolve at similar rates. Nature **407**:900–903.

WOESE, C. R., O. KANDLER, and M. L. WHEELIS. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA **87**:4576–4579.

WOLFE, K. H., P. M. SHARP, and W. H. LI. 1989. Mutation rates differ among regions of the mammalian genome. Nature **337**:283–285.

YANG, H., S. FITZ-GIBBON, E. M. MARCOTTE, J. H. TAI, E. C. HYMAN, and J. H. MILLER. 2000. Characterization of a thermostable DNA glycosylase specific for U/G and T/G mismatches from the hyperthermophilic archaeon *Pyrobaculum aerophilum*. J. Bacteriol. **182**:1272–1279.

YANG, Y., A. L. JOHNSON, L. H. JOHNSTON, W. SIEDE, E. C. FRIEDBERG, K. RAMACHANDRAN, and B. A. KUNZ. 1996. A mutation in *Saccharomyces cerevisiae* gene *(RAD3)* required for nucleotide excision repair and transcription increases the efficiency of mismatch correction. Genetics **144**: 459–466.

YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**: 555–556.

YANG, Y., X. KANG, L. KOHALMI, R. KARTHIKEYAN, and B. A. KUNZ. 1999. Strand interruptions confer strand preference during intracellular correction of a plasmid-borne mismatch in *Saccharomyces cerevisiae*. Curr. Genet. **35**:499–505.

ZAR, J. H. 1984. Biostatistical analysis. Prentice Hall, Englewood Cliffs, NJ.

ZHU, B., Y. ZHENG, H. ANGLIKER, S. SCHWARZ, S. THIRY, M. SIEGMANN, and J. P. JOST. 2000. 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. Nucleic Acids Res. **28**:4157–4165.

KEN WOLFE, reviewing editor

Accepted March 12, 2002