

Letter to the Editor

GC-Content Evolution in Mammalian Genomes: The Biased Gene Conversion Hypothesis

N. Galtier,* G. Piganeau,[†] D. Mouchiroud[†] and L. Duret[†]

*UMR CNRS 5000, "Génome, Populations, Interactions," Université Montpellier 2, 34095 Montpellier Cedex, France and [†]UMR CNRS 5558, "Biométrie et Biologie Evolutive," Université Claude Bernard, Lyon 1, 69622 Villeurbanne, France

Manuscript received December 27, 2000

Accepted for publication July 31, 2001

BASE composition is spatially structured in mammalian genomes. From sodium chloride centrifugation experiments, BERNARDI *et al.* (1985) defined three major classes of genomic fragments with low, median, and high GC content, respectively, and called them isochores. This discrete description now appears artificial. Analyses of the complete human genome (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2001) have dismissed the underlying hypothetical picture of sharp boundaries between long homogeneous fragments: GC content turns out to vary continuously, and somewhat erratically, along chromosomes. These analyses, however, have confirmed the existence of a highly significant spatial autocorrelation of GC content, with most of the structure detectable at a relatively large (300-kb) scale. There is a strong correlation between the GC content at third codon positions of codons (GC3) and the GC content of the region in which a gene is located (BERNARDI 2000). GC3 varies from a typical 40% in low-GC-content regions to 80% and more in high-GC-content regions. GC content is correlated with various genomic features, including repeat element distribution, methylation pattern (JABBARI and BERNARDI 1998), and, most remarkably, gene density (MOUCHIROUD *et al.* 1991; DURET *et al.* 1995). GC-rich regions include many genes with short introns while GC-poor regions are essentially deserts of genes. This suggests that the distribution of GC content in mammals could have some functional relevance, raising the issue of its origin and evolution. For brevity, in this article, we use the word "isochore" as an abbreviation for "the peculiar, structured distribution of GC content in mammalian genomes."

Two opposite views about isochore evolution have been hotly debated over the last 20 years, as part of the neutralist/selectionist controversy. One view is that

isochores may simply reflect variable mutation processes among genomic regions, consistent with the neutral model. Such variation in mutational biases has never been demonstrated. Alternatively, isochores might be the result of natural selection. Bernardi and colleagues have long argued that the isochore structure is an adaptation to homeothermy since it has been found in mammals and birds but not amphibians and fish (*e.g.*, BERNARDI 1993). The discovery of isochores in crocodiles and turtles (HUGHES *et al.* 1999) led to the rejection of this hypothesis. Moreover, selection, if any, must be unrelated to gene expression: The GC content of genes in humans does not correlate positively with their expression level or pattern (GONCALVES *et al.* 2000), and even pseudogenes translocated into a GC-rich region undergo an increase of GC content (FRANCINO and OCHMAN 1999). Neutralists furthermore argued that, given the small effective population size of mammalian species, a selective hypothesis must imply very high selection coefficients (typically 2 orders of magnitude higher than in *Drosophila*) at every position of GC-rich regions, including introns and intergenic DNA (SHARP *et al.* 1995). The existence of such a selective pressure without apparent correlation with gene expression appeared quite speculative.

On the other hand, recent analyses of human polymorphism data sets (EYRE-WALKER 1999; SMITH and EYRE-WALKER 2001) unexpectedly contradicted the mutational bias hypothesis. Under neutrality, the substitution process (accumulation of changes in the long run) should reflect the mutation process since each kind of mutation has an equal fixation probability. At equilibrium, therefore, the number of AT → GC and GC → AT mutations arising should be equal. Eyre-Walker, however, found significantly more GC → AT than AT → GC mutations. G and C alleles seem to have a selective advantage, making the substitution process more GC biased than the mutation process. These reports renewed interest in selective hypotheses for the evolution of isochores (*e.g.*, BERNARDI 2000). There are, however, at least two alternative possible explanations for this

Corresponding author: N. Galtier, UMR 5000, "Génome, Populations, Interactions," Université Montpellier 2, CC 63, Place E. Bataillon, 34095 Montpellier Cedex, France. E-mail: galtier@univ-montp2.fr

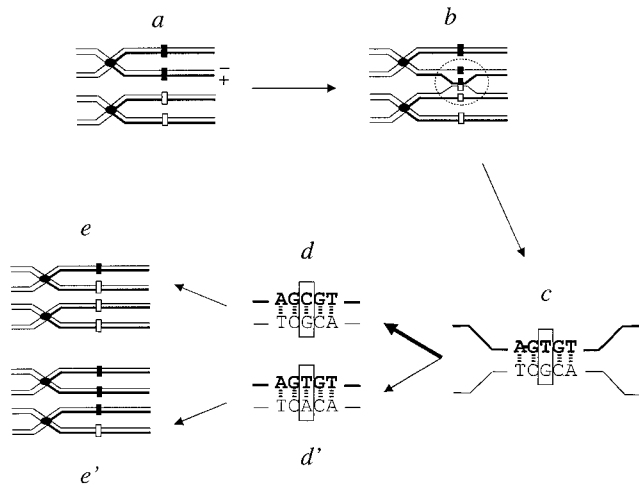


FIGURE 1.—Gene conversion during a recombination event. A bivalent with a heterozygous site is considered (a: solid rectangle, A:T pair; open rectangle, G:C pair). (b) A heteroduplex occurs during a recombination event. (c) The heterozygous site results in a T:G mismatch in the heteroduplex. (d) The T:G mismatch is repaired to C:G or (d') A:T, resulting in non-Mendelian segregation of alleles (e, e'). A bias occurs when repair to G:C is more (less) frequent than repair to A:T.

result. First, the excess of GC \rightarrow AT mutations may be the consequence of nonstationarity. Such an excess is expected if GC content is decreasing in GC-rich regions. SMITH and EYRE-WALKER (2001) could not detect any departure from stationarity by comparing human sequences to other primate species. An analysis of repeated elements in the human genome, however, somewhat contradicted this view (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2001). Second, HOLMQUIST (1992) and EYRE-WALKER (1993, 1999) mention another possible evolutionary force that might explain the polymorphism pattern, namely biased gene conversion (BGC). We now examine the BGC hypothesis and argue that it is likely to play an important role in the evolution of GC content in mammals.

Gene conversion is a molecular mechanism associated with recombination in which a genomic fragment is "copied/pasted" onto another homologous fragment (LAMB 1984). Both DNA fragments therefore share identical sequences after a conversion event. Allelic conversion occurs during the process of meiotic recombination. A DNA heteroduplex is formed, involving the plus strand of one chromosome and the minus strand of the sister chromosome (Figure 1). If this region of heteroduplex includes a heterozygous site, then a mismatch will occur. This mismatch may be recognized and corrected by DNA repair systems, thus removing a difference between the two chromosomes. Conversion results in non-Mendelian segregation of alleles in the germ cell where it occurs (Figure 1). This has no evolutionary relevance if the two possible ways of repairing a mismatch have equal probabilities—the proportions of ga-

metes averaged over germ cells remain Mendelian. If, however, the repair process is biased toward, say, G:C pairs, then allele frequencies would evolve in a nonneutral fashion: GC/AT heterozygotes would produce a higher proportion of GC gametes than AT gametes, resulting in a higher fixation probability for GC alleles.

NAGYLAKI (1983) showed that the dynamics of the fixation process for one locus under biased gene conversion (BGC) are identical to that under directional selection. The two effects could therefore account for the excess of GC \rightarrow AT polymorphisms in humans. EYRE-WALKER (1999) provides arguments against the BGC hypothesis. First, he notes that BGC is unlikely since third codon positions and introns have distinctive GC contents, while BGC should have identical effects in intron and exons. This difference, however, can be explained by an external factor, namely the accumulation of relatively GC-poor transposons (LINES, Alu) within introns but not within coding regions (DURET and HURST 2001). There is therefore no need to invoke distinct fixation biases in introns and synonymous sites. EYRE-WALKER (1999) also argues that the BGC hypothesis is unlikely because there is only a narrow range (one order of magnitude) within which the rate of biased conversion would be high enough to significantly alter polymorphism patterns but low enough not to induce an extreme base composition. Biased conversion within the necessary range is not just an *ad hoc* assumption, however, if one takes into account the selective pressure acting on genetic systems. Very highly biased conversion rates would probably simply be selected against (see below for additional discussion of this issue).

Several observations indicate that the BGC hypothesis deserves further attention. First, biased DNA repair toward GC has been observed experimentally in mammalian cells after transfection of mismatched DNA fragments (BROWN and JURICNY 1987; BILL *et al.* 1998). To significantly influence the fixation process of GC/AT polymorphisms and the resulting equilibrium GC content, BGC must occur at a rate w such that $N_e \cdot w \geq 1$ (where N_e is the effective population size and w is the per nucleotide conversion rate times the average conversion bias). Assuming effective population sizes of $\sim 10^4$ – 10^5 , the estimated recombination rates ($\sim 10^{-8}$ crossing over per base pair per generation; WEISSENBACH *et al.* 1992), conversion biases (BILL *et al.* 1998), and length of heteroduplex during recombination events (several hundreds of bases and up to several kilobases; DETLOFF and PETES 1992; LI and BAKER 2000) are high enough in mammals for BGC to be effective.

Sequence analysis also provides clues about suggestions of a putative role of gene conversion. Two kinds of conversion events occur within genomes: conversion between different copies within a gene family (ectopic gene conversion) or conversion between the two alleles of a gene (allelic gene conversion, Figure 1). If BGC were a major determinant of GC-content evolution, one

TABLE 1

GC content of ribosomal and transfer RNA in humans

Gene	Length	GC content (%)
5S rRNA	121	59.5
5' ETS	3656	79.2
18S rRNA	1871	56.1
ITS1	1095	79.6
5.8S rRNA	157	57.3
ITS2	1155	83.1
28S rRNA	5035	69.2
3' ETS	381	81.9
tRNA	65-88	58 ^a

Data from GenBank (accession nos. U13369 and X71799). Note that the average GC content in the human genome is 41%.

^a Average over 46 nuclear tRNA sequences.

would expect sequences undergoing frequent gene conversion—either ectopic or allelic—to become GC rich. Ectopic conversion is frequent in genes undergoing concerted evolution, and allelic conversion is frequent in recombination hotspots. Among the genes undergoing concerted evolution in mammals, the well-known ribosomal operons (especially their introns, Table 1), transfer RNAs (Table 1), and histones (GC3 range from ~60% to ~90%) are all GC rich, consistent with the above prediction. In histones the GC content of coding sequences (which experience gene conversion) is much higher than expected given the GC content of flanking regions (which do not; DE BRY and MARZLUFF 1994). Similarly, HOGSTRAND and BÖHME (1999) reported that those regions of the human and mouse MHC genes involved in gene conversion events show a higher CpG level and a higher GC content than regions for which conversion is not suspected.

Hotspots of recombination should also become GC rich under the BGC hypothesis, since gene conversion between alleles occurs during meiotic recombination, where sister chromatids are paired. EYRE-WALKER (1992) reported observations suggesting a correlation between recombination rates and GC contents, including a higher chiasma density in cytogenetic bands suspected to be GC richer and a low GC content of the nonrecombining Y chromosome. The most famous hotspot of recombination in mammals is the pseudoautosomal region (PAR) of the X and Y chromosomes. In the sexual chromosomes, a short region of homology—the so-called PAR—behaves like an autosomal bivalent in males. During meiosis, every bivalent undergoes at least one recombination event (*e.g.*, see LAWRIE *et al.* 1995). Given its small size, the per nucleotide recombination rate in the PAR is therefore much higher than in autosomes (SORIANO *et al.* 1987). X-specific chromosomal sequences, in contrast, undergo one-half the autosomal recombination rate, while Y-specific sequences do not

TABLE 2

Average GC content at third codon position (GC3) in genes located in autosomes, pseudoautosomal regions (PAR), or non-PAR regions of X and Y chromosomes

Chromosome	Average GC3	SE	Gene no.
Y	50.76	3.58	23
X	56.30	0.64	438
Autosome	62.14	0.21	5729
PAR	76.19	4.28	11

Complete human protein coding sequences were selected from HOVERGEN (DURET *et al.* 1994) release 40 (May 2000). Pseudogenes and sequences <100 bp were excluded. We also dismissed immunoglobulin and T-cell receptor genes. The average GC3 is higher in PAR than in autosomes (Student's *t*-test = 2.96, *P* = 0.003) and higher in autosomes than in X (Student's *t*-test = 7.57, *P* < 0.0001). The average GC3 is slightly higher in X than in Y, but the test is only marginally significant (Student's *t*-test = 1.89, *P* = 0.0593).

recombine at all. The average GC content at third codon positions of genes in these four chromosomal locations (Y, X, autosome, and PAR) increases with the recombination rate (Table 2), consistent with the BGC hypothesis. The GC content is even higher in the mouse PAR (gene STS: GC3 = 96.2%). PERRY and ASHWORTH (1999) showed that the GC3 of a gene recently translocated into the PAR in mouse increased from 50 to 73% in <1 million years, strongly suggesting that recombination is the cause, not the consequence, of a high GC content. A similar line of evidence comes from birds. Bird genomes include very small chromosomes. Like the PAR, these microchromosomes must have a high per nucleotide recombination rate since they undergo at least one recombination event per generation. Bird microchromosomes again show a very high GC content, probably even higher than mammalian GC-rich regions (KADI *et al.* 1993). A relationship between recombination rate and GC content in humans was recently demonstrated by EISENBARTH *et al.* (2000). These authors analyzed the linkage disequilibrium between loci at the boundary of a GC-rich and a GC-poor region. They found significantly greater linkage disequilibrium (*i.e.*, presumably less recombination) in the GC-poor region, again consistent with the BGC hypothesis. This result is somewhat supported by a genome-wide analysis: YU *et al.* (2001) reported a weak, though significant, correlation between estimated recombination rates and GC content, despite the low level of accuracy of the human genetic map, as the authors acknowledge. Finally, note that a relationship between recombination and GC content has also been found in *Saccharomyces cerevisiae* (GERTON *et al.* 2000), *Drosophila melanogaster*, and *Caenorhabditis elegans* (MARAIS *et al.* 2001).

The BGC hypothesis can therefore account for (i) the higher GC content of regions undergoing gene conversion; (ii) the correlation between recombination rate

and GC content, where recombination seems to be the governing force; and (iii) the nonneutral human polymorphism patterns. These arguments, together with the experimental evidence of a GC bias of the repair process, strongly suggest (but do not demonstrate) that BGC might be a major force governing isochore evolution.

A correlation between GC content and recombination rate is also expected under the hypothesis that sequences are under selection, because linkage reduces the efficacy of selection (HILL and ROBERTSON 1966). This hypothesis, however, requires the existence of high selection coefficients, which are unable to operate in regions of low recombination rate. GC3 in the mouse PAR is >95%, which is reached for $N_e \cdot s = 1$ assuming that sites evolve independently (where N_e is the effective population size and s the per site selection coefficient; *e.g.*, see BULMER 1991). If this were true, the mutation load for the entire genome would be enormous. If Hill-Robertson effects were impeding such a strong selection pressure and reducing the genomic GC content to a value so far from its optimal value (the GC3 in GC-poor regions is typically 40%), one might expect natural selection to favor modifiers, increasing the recombination rate (although this has not been demonstrated formally). Furthermore, such a strong Hill-Robertson effect, if any, should not affect GC content only. Under this hypothesis, selection against deleterious nonsynonymous changes should be less effective in regions of low recombination rate. This should result in a negative correlation between GC content and nonsynonymous substitution rates. We examined this prediction and found a weak positive ($r = 0.036$) correlation between a maximum-likelihood estimate of the nonsynonymous rate (YANG and NIELSEN 2000) and the GC content of 5000 genes sampled in human and mouse (not shown). The Hill-Robertson effect, if any, has no detectable effect on the fate of nonsynonymous changes, making the hypothesis that it influences putatively selected synonymous ones unlikely.

A test of the causal relationship between recombination and GC content could come from the examination of polymorphism patterns in inbreeding species. Inbreeding populations include no heterozygotes so that BGC, however strong, would not influence the dynamics of allele segregation. Inbreeding species are scarce in vertebrates but common in plants. The genomes of Gramineae are quite similar to mammalian genomes with respect to the distribution of GC content and genes (CARELS and BERNARDI 2000). A comparison of sequence polymorphism patterns in inbreeding and outbreeding species of Gramineae and maybe other plant species would be of great relevance to the study of GC content evolution.

Although, under BGC, alleles do not segregate according to the neutral model, BGC is basically a neutral process since there are no fitness differences between

individuals. The BGC hypothesis, however, might ironically raise new selectionist issues. Even if direct selection for GC content at every site of GC-rich regions turns out to be unlikely, selection on the molecular machinery that determines the dynamics of GC content must be considered. An interesting hypothesis was proposed by FRYXELL and ZUCKERKANDL (2000). They argued that a biased repair process might be an adaptation to the high rate of methyl-cytosine deamination. Cytosines involved in CpG doublets have a mutation rate maybe 10 times higher than other nucleotides in humans (GIANNELLI *et al.* 1999). Correctly repairing these mutations is therefore crucial. Deamination of unmethylated cytosines produces uracil, and the resulting U:G mismatch can easily be repaired since U is an alien base in DNA. Deamination of methyl-cytosine, however, produces thymine and an ambiguous T:G mismatch. A repair process that would favor T:G \rightarrow C:G repair over T:G \rightarrow T:A might therefore be advantageous in methylated genomes (including mammals, birds, and plants) since most T:G mismatches result from a C \rightarrow T mutation. Under this scenario (and assuming again that BGC determines GC content), isochores would be a by-product of natural selection acting on DNA repair. A potential test of this would be to compare mammals and species in which the repair of G:T mismatches is unbiased to see whether the latter have similarly structured CG content. Few such data are available, but, consistently, one such species, *Xenopus laevis* (VARLET *et al.* 1990), has a genome with much less structured GC content than that of the mammals that have been studied.

LITERATURE CITED

- BERNARDI, G., 1993 The vertebrate genome: isochores and evolution. *Mol. Biol. Evol.* **10**: 186–204.
- BERNARDI, G., 2000 Isochores and the evolutionary genomics of vertebrates. *Gene* **241**: 3–17.
- BERNARDI, G., B. OLOFSSON, J. FILIPSKI, M. ZERIAL, J. SALINAS, *et al.*, 1985 The mosaic genome of warm-blooded vertebrates. *Science* **228**: 953–958.
- 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.
- BROWN, T. C., and J. JIRICNY, 1987 A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* **50**: 945–950.
- BULMER, M., 1991 The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**: 897–907.
- CARELS, N., and G. BERNARDI, 2000 Two classes of genes in plants. *Genetics* **154**: 1819–1825.
- DE BRY, R. W., and W. F. MARZLUFF, 1994 Selection on silent sites in the rodent H3 histone gene family. *Genetics* **138**: 191–202.
- DETLOFF, P., and T. D. PETES, 1992 Measurements of excision repair tracts formed during meiotic recombination in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**: 1805–1814.
- DURET, L., and L. D. HURST, 2001 The elevated G and C content at exonic third sites is not evidence against neutralist models of isochore evolution. *Mol. Biol. Evol.* **18**: 757–762.
- DURET, L., D. MOUCHIROUD and M. GOUY, 1994 HOVERGEN—a database of homologous vertebrate genes. *Nucleic Acids Res.* **22**: 2360–2365.

- DURET, L., D. MOUCHIROUD and C. GAUTIER, 1995 Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. *J. Mol. Evol.* **40**: 308–317.
- 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.
- 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.
- EYRE-WALKER, A., 1993 Recombination and mammalian genome evolution. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **252**: 237–243.
- EYRE-WALKER, A., 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.
- FRANCINO, M. P., and H. OCHMAN, 1999 Isochores result from mutation, not selection. *Nature* **400**: 30–31.
- FRYXELL, K. J., and E. ZUCKERKANDL, 2000 Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol. Biol. Evol.* **17**: 1371–1383.
- GERTON, J. L., J. DE RISI, R. SHROFF, M. LICHTEN, P. O. BROWN *et al.*, 2000 Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **97**: 11383–11390.
- GIANELLI, F., T. ANAGNOSTOPOULOS and P. M. GREEN, 1999 Mutation rates in human. II. Sporadic mutation-specific rates and rate of detrimental human mutations inferred from Hemophilia B. *Am. J. Hum. Genet.* **65**: 1580–1587.
- GONCALVES, I., L. DURET and D. MOUCHIROUD, 2000 Nature and structure of human genes that generate retropseudogenes. *Genome Res.* **10**: 672–678.
- HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. *Genet. Res.* **7**: 44–57.
- HOGSTRAND, K., and J. BOHME, 1999 Gene conversion of major histocompatibility complex genes is associated with CpG-rich regions. *Immunogenetics* **49**: 446–455.
- HOLMQUIST, G. P., 1992 Chromosome bands, their chromatine flavours, and their functional features. *Am. J. Hum. Genet.* **51**: 17–37.
- HUGHES, S., D. ZELUS and D. MOUCHIROUD, 1999 Warm-blooded isochore structure in Nile crocodile and turtle. *Mol. Biol. Evol.* **16**: 1521–1527.
- INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001 Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- JABBARI, K., and G. BERNARDI, 1998 CpG doublets, CpG islands and Alu repeat elements in long human DNA sequences from different isochores families. *Gene* **224**: 123–128.
- KADI, F., D. MOUCHIROUD, G. SABEUR and G. BERNARDI, 1993 The compositional patterns of the avian genomes and their evolutionary implications. *J. Mol. Evol.* **37**: 544–551.
- LAMB, B. C., 1984 The properties of meiotic gene conversion important in its effects on evolution. *Heredity* **53**: 113–138.
- LAWRIE, N. M., C. TEASE and M. A. HULTEN, 1995 Chiasma frequency, distribution and interference maps of mouse autosomes. *Chromosoma* **104**: 308–314.
- LI, J., and M. D. BAKER, 2000 Use of a small palindrome genetic marker to investigate mechanisms of double-strand-break repair in mammalian cells. *Genetics* **154**: 1281–1289.
- 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.
- MOUCHIROUD, D., G. D'ONOFRIO, B. AISSANI, G. MACAYA, C. GAUTIER *et al.*, 1991 The distribution of genes in the human genome. *Gene* **100**: 181–187.
- NAGYLAZI, T., 1983 Evolution of a finite population under gene conversion. *Proc. Natl. Acad. Sci. USA* **80**: 6278–6281.
- PERRY, J., and A. ASHWORTH, 1999 Evolutionary rate of a gene affected by chromosomal position. *Curr. Biol.* **9**: 987–989.
- SHARP, P. M., M. AVEROF, A. T. LLOYD, G. MATASSI and J. F. PEDEN, 1995 DNA sequence evolution: the sounds of silence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **349**: 241–247.
- SMITH, N. G. C., and A. EYRE-WALKER, 2001 Synonymous codon bias is not caused by mutation bias in human. *Mol. Biol. Evol.* **18**: 982–986.
- SORIANO, P., E. A. KEITGES, D. F. SCHORDERET, K. HARBERS, S. M. GARTLER *et al.*, 1987 High-rate of recombination and double crossovers in the mouse pseudoautosomal region during male meiosis. *Proc. Natl. Acad. Sci. USA* **84**: 7218–7220.
- VARLET, I., M. RADMAN and P. BROOKS, 1990 DNA mismatch repair in *Xenopus* egg extracts: repair efficiency and DNA repair synthesis for all single base-pair mismatches. *Proc. Natl. Acad. Sci. USA* **87**: 7883–7887.
- WEISSENBACH, J., G. GYAPAY, C. DIB, A. VIGNAL, J. MORISSETTE *et al.*, 1992 A second-generation linkage map of the human genome. *Nature* **359**: 794–801.
- YANG, Z. H., and R. NIELSEN, 2000 Estimating synonymous and non-synonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**: 32–43.
- YU, A., C. ZHAO, Y. FAN, W. JANG, A. J. MUNGALL *et al.*, 2001 Comparison of human genetic and sequence-based physical maps. *Nature* **409**: 951–953.

Communicating editor: D. CHARLESWORTH