Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes

Laurent Duret[1,*] and Nicolas Galtier[2]

[1]Université de Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Évolutive, F-69622, Villeurbanne, France; email: duret@biomserv.univ-lyon1.fr

[2]Université Montpellier 2, CNRS UMR5554, Institut des Sciences de l'Evolution, 34095 Montpellier, France; email: nicolas.galtier@univ-montp2.fr

Annu. Rev. Genomics Hum. Genet. 2009. 10:285–311

First published online as a Review in Advance on July 16, 2009

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi: 10.1146/annurev-genom-082908-150001

1527-8204/09/0922-0285$20.00

*To whom correspondence should be sent.

Key Words

isochores, recombination, selection, neutral evolution, mutation, comparative genomics

Abstract

Recombination is typically thought of as a symmetrical process resulting in large-scale reciprocal genetic exchanges between homologous chromosomes. Recombination events, however, are also accompanied by short-scale, unidirectional exchanges known as gene conversion in the neighborhood of the initiating double-strand break. A large body of evidence suggests that gene conversion is GC-biased in many eukaryotes, including mammals and human. AT/GC heterozygotes produce more GC- than AT-gametes, thus conferring a population advantage to GC-alleles in high-recombining regions. This apparently unimportant feature of our molecular machinery has major evolutionary consequences. Structurally, GC-biased gene conversion explains the spatial distribution of GC-content in mammalian genomes—the so-called isochore structure. Functionally, GC-biased gene conversion promotes the segregation and fixation of deleterious AT → GC mutations, thus increasing our genomic mutation load. Here we review the recent evidence for a GC-biased gene conversion process in mammals, and its consequences for genomic landscapes, molecular evolution, and human functional genomics.

INTRODUCTION

Crossover: recombination event leading to the reciprocal exchange of flanking markers

Noncrossover: recombination event without exchange of flanking markers

Gene conversion: nonreciprocal copying of a stretch of DNA from one chromosome onto the other chromosome

BGC: biased gene conversion

DSB: double-strand break

In mammals, as in most eukaryotes, homologous chromosomes exchange genetic information through recombination during meiosis. Meiotic recombination is considered to be a fundamental process for two reasons: first, crossovers are required for the proper segregation of chromosomes during meiosis; second, crossovers create new combinations of alleles, and hence contribute to increasing the genetic diversity on which selection can act (see Reference 28 for a review). The consequences of meiotic recombination are, however, not limited to crossovers. Indeed, a fraction of recombination events are resolved as noncrossovers, and in both crossovers and noncrossovers, recombination involves gene conversion, i.e., the nonreciprocal copying of a stretch of DNA from one chromosome onto the other chromosome.

A meiotic gene conversion event leads to non-Mendelian segregation of gametes derived from the germ cell where it occurs. This has no consequences at the population level if both alleles have the same probability of conversion, because the whole gamete pool (averaged over germ cells) remains unbiased (i.e., 50:50 ratio of both alleles). If, however, one allele has a higher probability to be the donor during a conversion event, then the pool of gametes produced by heterozygote individuals will contain a higher frequency of this allele. Such a process of biased gene conversion (BGC) confers on the donor allele a higher probability of transmission to the next generation, and therefore an evolutionary advantage over the acceptor allele. Many examples of BGC have been reported in different eukaryotic taxa (85). Over 20 years ago, several theoretical studies were initiated to investigate the potential consequences of this process on the evolution of gene sequences (e.g., 8, 103). However, the consequences of gene conversion on genome evolution have been largely unexamined until recently.

Here we review recent evidence that BGC has had a major impact on the evolution of mammalian genomic landscapes. This BGC process, which is probably widespread in eukaryotes, affects not only neutral sites (noncoding regions, synonymous sites) but also functional sites (coding regions, regulatory elements), and hence can confound classical tests used to detect selective effects. Both theoretical development and empirical data indicate that BGC can lead to the fixation of deleterious mutations, and hence that recombination hotspots might constitute the Achilles' heel of mammalian genomes.

GENE CONVERSION: THE MECHANISMS

Recombination (Crossovers and Noncrossovers) Involves Gene Conversion

Current knowledge of meiotic recombination derives mainly from studies in yeast. But the strong conservation of the major proteins involved indicates that recombination processes should be similar in yeasts and in mammals (31). Meiotic recombination is initiated by the formation of a double-strand break (DSB) made by the Spo11 protein. The DSB ends are then degraded to generate long 3′ single-strand tails of several hundred base pairs. One single-stranded DNA invades the homologous sequence on the other (uncut) chromosome (**Figure 1**). This intermediate can then be repaired via different pathways that have two possible outcomes: crossovers and noncrossovers (**Figure 1**) (for more details on the current model of recombination, see References 26, 31, 140 for recent reviews). In all cases, recombination involves the formation of heteroduplex DNA, at various steps of the process (shadowed rectangles in **Figure 1**). When the two parental alleles are not identical, the resulting mismatches in the heteroduplex are generally repaired. Depending on the choice of the corrected strand, this repair can lead to gene conversion or restoration. The different possible outcomes of DNA repair in the gene conversion tract, and the corresponding segregation ratio are presented in **Figure 2**.

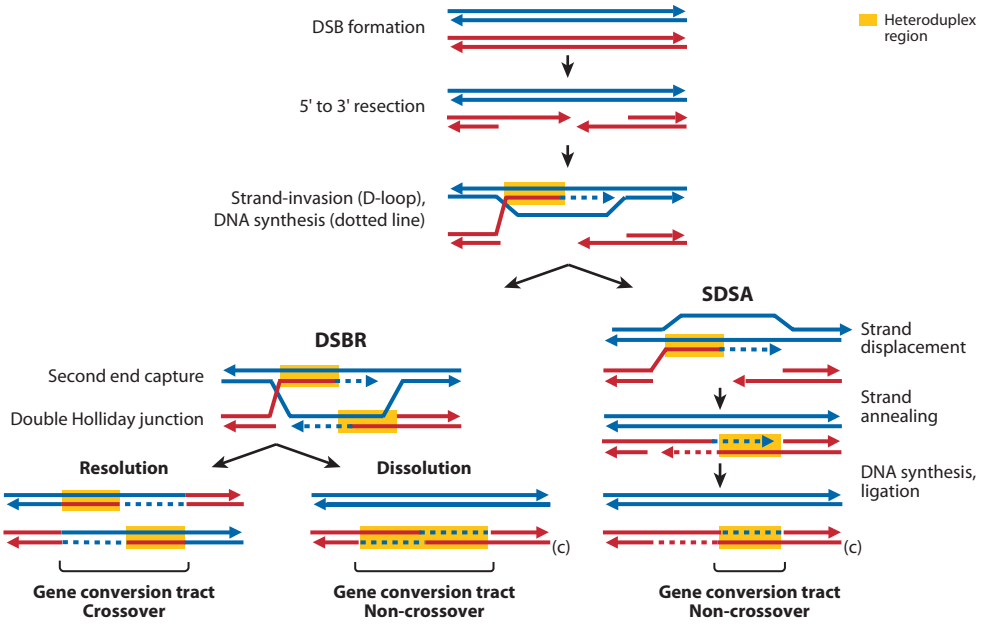

Figure 1

Mechanisms of recombination and gene conversion. Meiotic recombination can lead to crossovers or noncrossovers. Three models have been proposed to explain these different pathways: double-strand-break repair (DSBR), dissolution of the double Holliday junctions and synthesis-dependent strand-annealing (SDSA) [for more details of these models, see Chen et al. (26)]. In all models, recombination is initiated by the formation of a double-strand-break (DSB), followed by 5′ to 3′ resection leading to 3′ ssDNA tails. One tail invades the homologous DNA duplex, forming a displacement (D)-loop, which is then extended by DNA synthesis. In the SDSA pathway, the invading strand (including newly synthesized DNA) is displaced from the template and anneals to the other 3′ end of the DSB, which leads to noncrossovers. In the DSBR pathway, the D-loop captures the second 3′ end. DNA synthesis and ligation lead to the formation of double Holliday junctions (dHJs), with heteroduplex DNA flanking the DSB site. In the canonical DSBR model, the resolution of this recombination intermediate (by DNA cleavage and ligation) is predicted to generate an equal number of crossovers and noncrossovers (not shown). However, it is established in *S. cerevisiae* that noncrossovers are not generated by this pathway. The dHJ recombination intermediate can also be removed by an alternative pathway (dissolution) that leads exclusively to noncrossovers. Heteroduplex DNA (indicated by *yellow rectangles*) is formed at various steps of these different pathways.

The prominent repair pathway during recombination is the mismatch repair (MMR) system, which is also involved in the repair of base misincorporations during DNA replication (128). The directionality of MMR is determined by the presence of DNA nicks: the strand containing a nick is degraded and resynthesized using the intact strand as a template. This implies that gene conversion occurs mainly from the uncut chromosome toward the chromosome that experienced the DSB. In agreement with the predictions of this model, analyses of meiosis products in yeast and humans indicated that in most cases, recombination leads to simple gene conversion tracks (**Figure 2**, Box 1). However, studies in MMR-deficient yeasts revealed that some other pathways are also involved in the repair of mismatches in heteroduplex DNA (128), which could contribute to the formation of complex conversion tracks (**Figure 2**).

Biased Gene Conversion

Biased gene conversion (BGC) occurs in humans. Among the 36 crossover hotspots that

MMR: mismatch repair

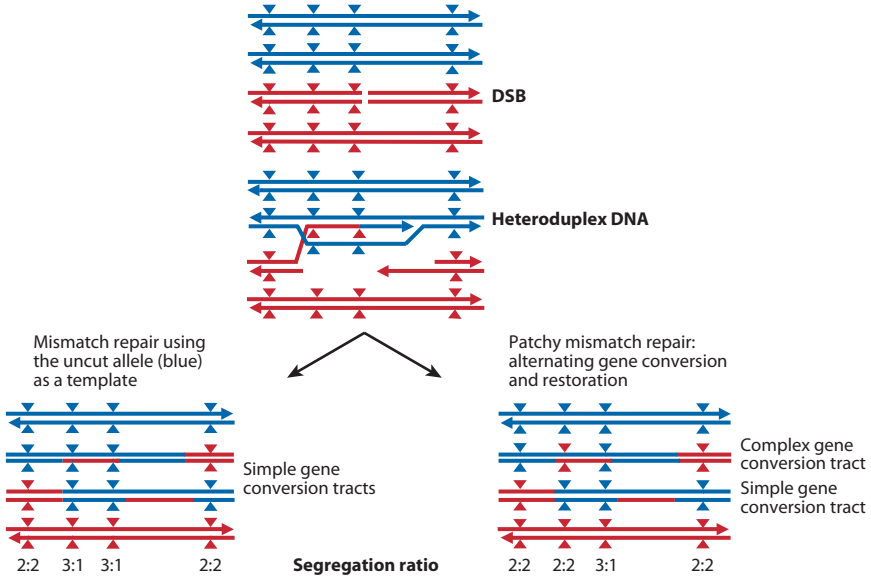

Figure 2

Examples of possible outcomes of mismatch repair in heteroduplex DNA. The example shown corresponds to a recombination event leading to a crossover, but the same patterns are expected for noncrossovers. Four chromatids are present at the time of initiation of recombination. The two different homologs are indicated in blue and red. Blue and red triangles indicate the position of genetic markers that differ between the two haplotypes. Pairs of blue and red triangles in heteroduplex DNA correspond to mismatches. The arrowheads indicate the 3′ ends of DNA. In most cases, recombination leads to simple gene conversion tracts (see Box 1). This results from the fact that mismatch repair is generally initiated by degrading the DNA strand containing a nick. Hence, the uncut allele (*blue*) is generally used as a template to repair mismatches. However, a fraction of recombination events lead to complex conversion tracts, alternating gene conversion and restoration. The segregation ratios in resulting gametes are indicated.

have been analyzed by sperm typing, at least 8 show distorted segregation ratios: 4 show strong transmission distortion (from 74:26 to 83:17) and 4 others show more subtle, but significant, distortion (69, 70, 134). Given that gene conversion occurs primarily from the uncut to the cut chromosome (Box 1), a conversion bias is expected when one haplotype is more prone to the formation of DSBs (the recombination initiation bias model). As predicted by this model, the hotspots with strong transmission distortion show recombination activity polymorphism between men: The most frequently converted allele corresponds to the haplotype with highest recombination activity (69, 70, 134). This conversion bias favoring recombination-suppressor haplotypes raises the question of how recombination hotspots are maintained in the genome (for discussions about this so-called hotspot paradox see 16, 27, 69). In addition to this recombination initiation bias, biased gene conversion might also result from a bias in the repair of mismatches in heteroduplex DNA. Such a DNA-repair bias is expected to be associated to complex gene conversion tracks (Box 1, **Figure 2**).

gBGC: GC-biased gene conversion

Experimental Evidence for Biased Gene Conversion Favoring GC-Alleles Over AT-Alleles

Genome-wide analyses of meiosis products uncovered gene conversion biases in yeast (92). Furthermore, BGC tends to favor GC-alleles over AT-alleles (15, 92). This process of GC-biased gene conversion (gBGC) is relatively weak: Among GC/AT heterozygote sites involved in recombination events, the GC-allele

is the donor in 50.62% of cases (significantly different from the 50% expected in the absence of bias: $\chi^2 = 10.9$, $p < 10^{-3}$). The strength of this GC-bias is the same in crossovers and noncrossovers [Note that these figures were computed using genotyping data described in (92), and kindly provided by R. Bourgon]. This GC-bias is relatively modest, but such a weak meiotic drive can strongly increase the probability of fixation of GC-alleles: What matters is the product of gBGC rate by effective population size (see below and Box 2). As expected, recombination rate and GC-content are positively correlated in yeast chromosomes (15, 17). In yeast, the repair of DNA mismatches in mitotic cells is GC-biased (15). It is therefore possible that the gBGC observed in meiosis is the direct consequence of this repair bias (15). Alternatively, a small increase of Spo11 activity due to an increase in AT content could bias the conversion. Although the rules influencing Spo11 action are unknown, Spo11 cleavages are not random and Spo11 is influenced by the DNA sequence context (31a).

In mammals, the repair of DNA mismatches in mitotic cells is strongly GC-biased (19, 20). This GC-biased repair process is probably an adaptation to limit the deleterious consequences of the hypermutability of methylated cytosines in these genomes (18). The GC-bias results from the activity of the base excision repair (BER) pathway, which involves DNA glycosylases that specifically excise thymines (and/or uracils) in DNA mismatches. If BER contributes to the repair of mismatches during recombination, then one would expect mammals to be subject to gBGC (20). There is considerable indirect evidence for gBGC in mammals (see below), but no direct evidence as yet. BER is known to be active in mammalian germ cells (108), but whether it contributes to the repair of mismatches in the DNA heteroduplex formed during recombination (where the main repair pathway is MMR) has not been established (21). In human, among the 8 known cases of BGC involving a GC/AT polymorphism, 4 favor the GC-allele and 4 the AT-allele (A. Jeffreys, personal communication; data from 69, 70, 134). The absence of GC-bias in these data is not surprising, given that many of these cases of biased gene conversion probably result from a recombination-initiation bias, and not from a DNA repair bias (134). In fact, much more data would be needed to detect a weak gBGC, such as that observed in yeast. Unfortunately, current techniques for the analysis of meiosis products in mammals are very tedious, and hence not amenable to scale up.

BER: base excision repair

Nonallelic Gene Conversion and Mitotic Gene Conversion

Homologous recombination is not restricted to meiosis. Indeed, one essential role of recombination is to repair DSBs caused by ionizing radiation, chemical agents or replication errors in mitotic cells (75). These DSBs are repaired preferentially using sister chromatids (not the homologs) (74), and hence generally do not lead to allelic gene conversion. However, some cases of mitotic gene conversion have been reported in human, which indicates that mitotic recombination occasionally involves homologs (71, 115). If mitotic gene conversion occurs at a substantial rate in the germline, it might also contribute to BGC. Finally, mitotic (and meiotic) recombination also occurs between homologous sequences present at different loci (i.e., ectopic or nonallelic recombination). This process leads to gene conversion between different copies of repeated sequences, and is notably responsible for the concerted evolution of gene families. There is indirect evidence that nonallelic gene conversion is also GC-biased in vertebrates (5, 49, 81) (see below).

THE IMPACT OF GC-BIASED GENE CONVERSION ON LARGE-SCALE HUMAN GENOMIC PATTERNS: THE HISTORY OF ISOCHORES

Investigatations into the impact of BGC on genome evolution were initiated by us and others in order to understand the evolution of a peculiar feature of mammalian genomic

BOX 1: GENE CONVERSION ASSOCIATED TO MEIOTIC RECOMBINATION IN YEAST AND HUMAN

In yeast it is possible to examine all four products of a given meiosis and to analyze them at very high spatial resolution (<100 bp) (92). The genome-wide mapping of crossovers and noncrossovers revealed that the median size of conversion tracts was 2 kb for crossovers and 1.8 kb for noncrossovers (the difference is statistically significant). Gene conversion affects about 1% of the genome at each meiosis (92). These conversion events are not uniformly distributed along yeast chromosomes: Both crossovers and noncrossovers occur primarily within recombination hotspots, which are preferentially located at the 5′ end of genes (overlapping the promoter and the coding region) (92). The crossover/noncrossover ratio is not constant across the genome: Some recombination hotspots favor crossovers, others noncrossovers (92).

The genotyping of recombination products in yeast showed that ~90% of crossovers are associated with simple conversion tracts. In most cases, the uncut chromosome is used as a template to repair mismatches in heteroduplex DNA (**Figure 2**), probably because of the preferential removal of strands with an available 3′ end. Thus, the direction of gene conversion is primarily determined by the choice of the cut allele: The chromosome where the DSB occurred will be the acceptor and the uncut chromosome will be the donor. However, a significant fraction (~10%) of crossovers are associated with complex, discontinuous conversion tracts, presumably arising through patchy repair of heteroduplex DNA at sites away from the DSB site (92).

In humans, the resolution of genetic maps based on the analysis of meiosis products in nuclear families is relatively limited [~100 kb for the most precise one (29)], but high-resolution genome-wide patterns of crossover events can be inferred indirectly by the analysis of allelic associations among populations. Such genetic maps reflect the sex-averaged crossover activity over several thousand generations (i.e., the historical crossover rate) (101). As in yeast, recombination events are not uniformly distributed: crossovers occur primarily in ~50,000 hotspots, typically less than 2 kb long (101). In humans, there is a marked decrease in crossover rate within genes (29, 101). However, like in yeast, the transcription start site corresponds to a local peak of recombination (46a).

(*Continued*)

landscapes, the so-called isochores. Over 30 years ago, Giorgio Bernardi's group demonstrated that the base composition of mammalian genomes is not homogeneous: These genomes show dramatic variations in GC-content over large scales (>100 kb) (13, 45). Bernardi initially proposed a model describing mammalian chromosomes as mosaics of discrete regions of relatively homogeneous base composition (the isochores) (13). With the advent of complete genome sequences, this model was criticized because it does not perfectly fit with the observed pattern: GC-rich regions show smaller-scale heterogeneities in base composition (105) and clear isochore boundaries cannot always be identified (86). However, the analysis of genome sequences fully confirmed the existence of large-scale variations in GC-content (86), and the term isochore, albeit not perfect, remains useful to describe this genomic organization (see Reference 42 for a review). These variations affect all genomic compartments—not only silent sites, which are presumably neutral or under very weak selective pressures, but also functional sites such as nonsynonymous codon positions (1, 13, 30, 99) (**Figure 3**). Gene density is much higher in GC-rich than in GC-poor isochores (38, 86, 98) (**Figure 3**). Moreover, this isochore organization is correlated with several other important genomic features, such as the timing of DNA replication (43, 133), the distribution of transposable elements (86, 125), and the rate of recombination (48). Thus, isochores clearly reflect fundamental aspects of mammalian genome evolution: What then is their origin? Did GC-rich isochores evolve because they confer some selective advantage, or do they simply result from nonadaptive evolutionary processes?

The Evolution of Isochores in the Amniote Lineage

The origin of mammalian isochores can only be understood by dating the evolution of these genomic landscapes. The base composition of homologous genomic regions is strongly

correlated between amniote species (mammals, birds, and reptiles) (24, 65, 73, 83), but no such isochore organization is observed in amphibians or fishes (12). Hence, the most parsimonious hypothesis is that the origin of mammalian isochores traces back to the last common ancestor of amniotes, 310 to 350 Mya (65, 83). This isochore organization is still evolving: in many, but not all, mammals, GC-rich isochores are eroding, i.e., their GC-content is decreasing (3, 4, 6, 34, 39, 86, 95, 138).

The Origin of Isochores: Selection or Mutation?

Several authors have proposed that GC-rich isochores might result from selection (10, 11, 41, 88, 130, 131). One difficulty with this model is that there is at present no satisfactory explanation for the putative selective advantage conferred by a higher GC-content. First, the evolutionary process responsible for the evolution of isochores affects not only genes, but also all kinds of noncoding sequences: introns, intergenic regions, pseudogenes, and transposable elements (3, 4, 35, 76, 95, 139). Thus, selection, if it is at work, is not acting primarily on the information content of genomic sequences, but simply on their GC-content. It has been suggested that GC-rich isochores have evolved in warm-blooded vertebrates (birds, mammals) to increase DNA stability (10). However, this model does not explain why only a fraction of the genome increased in GC-content. Moreover, there is no correlation between GC-content and body temperature in vertebrates (7, 119), and there is evidence that GC-rich isochores evolved in the last common ancestor of amniotes, before the origin of homeothermy in mammalian and bird lineages (65, 83).

It has also been proposed that the isochore organization of mammalian genomes might reflect an adaptation for fine-tuning the expression of tissue-specific genes (88, 130, 131). Directed mutagenesis experiments have shown that radical changes in GC-content in exons strongly affect gene expression levels in mammalian cell cultures (82). However, the correlations observed in mammalian genomes between gene expression patterns or levels and GC-content are extremely weak (121, 122). Thus, although the regional GC-content might affect gene expression, this feature does not appear to be the target of selection for regulating expression. In fact, this could have been anticipated: Changing the GC-content of a gene (typically 10 to 50 kb long) requires a very large number of substitutions. Conversely, a few mutations within regulatory elements can have a strong direct effect on gene expression. Thus, if there is a selective pressure to modify the expression level of a given gene, mutations occurring within regulatory elements will be selected well before any significant change in GC-content of the gene can occur.

The evolution of isochores results from the accumulation of base substitutions. Hence, if selectionist models are correct, then there must be a significant fitness difference between two individuals that differ by only one single point mutation in a >100-kb-long isochore. Even the

Direct analysis of recombination events in males at high resolution (<1 kb) is possible by sperm typing. Unfortunately, this approach is very labor intensive, and hence only 36 crossover hotspots have been analyzed to date in humans (134). These data confirmed that crossover hotspots predicted by population genetic analysis correspond to genuine crossover hotspots (although the actual hotspot intensity is not well predicted by the historical crossover rate) (134). The genome-wide distribution of noncrossover events in the human genome is not yet known. Sperm-typing analyses demonstrated that crossover hotspots generally also correspond to regions of high rates of noncrossover recombination (63, 68). However, as in yeast, the noncrossover/crossover ratio varies widely among recombination hotspots (from ~3:1 at hotspot *DNA3* to <1:12 at the β-globin hotspot) (63). The limited data presently available indicate that gene-conversion tracts are usually shorter than in yeast, of the order of 200 bp to 1 kb in length (26), and that noncrossovers lead to shorter conversion tracts than do crossovers (68). The great majority of crossovers are associated with simple conversion tracts, with only 0.33% of crossovers displaying complex haplotype switching (134). Thus, as in yeast, the direction of gene conversion is generally (but not always) from the uncut to the cut chromosome.

BOX 2: THE gBGC COEFFICIENT AND THE PROBABILITY OF FIXATION OF MUTATIONS AT NEUTRAL SITES

Consider a site with GC/AT polymorphism. The frequency of the GC-allele among gametes produced by a heterozygous individual is $x_{GC} = \frac{1}{2}(1+b)$, where b is the gBGC coefficient. Similarly, the frequency of the AT-allele in the pool of gametes is $x_{AT} = \frac{1}{2}(1-b)$.

BGC behaves just like selection of a semidominant mutation (103). So, at selectively neutral sites, the probability of fixation of AT → GC mutations is

$$P(AT \rightarrow GC) = \frac{1 - e^{-2b}}{1 - e^{-4Neb}}$$

and the probability of fixation of GC → AT mutations is

$$P(GC \rightarrow AT) = \frac{1 - e^{2b}}{1 - e^{4Neb}}$$

where N_e is the effective population size.

The gBGC coefficient at a given site depends on two parameters: (*i*) the frequency at which this site is involved in a recombination event (crossover or noncrossover) and (*ii*) the strength of the gene conversion bias. Under the assumption that gBGC is caused by the repair of DNA mismatches in heteroduplex DNA, the first parameter depends on the recombination rate and on the length of gene conversion tracts.

proponents of selective models admit that this is highly unlikely (11). Bernardi (11) proposed a "neoselectionist theory" to explain the evolution of isochores, but he did not present any analytical or simulation studies to test the validity of his model (11). In fact, theoretical studies showed that selection on large numbers of linked sites is inefficient, especially in species with relatively small effective population sizes such as mammals (112).

Because selection seems implausible, several authors proposed that variations of GC-content along mammalian chromosomes might result from variations in patterns of mutation (44, 46, 47, 142). But this model was rejected by the data. One strong prediction of such mutational models is that at neutral sites, all mutations should have the same probability of fixation. To test this prediction, Eyre-Walker (41) analyzed polymorphism at silent sites in the MHC locus in human and murine populations. He showed that in both species derived GC-alleles (i.e., alleles resulting from a AT → GC mutation) segregate at a higher frequency than derived AT-alleles. This observation, later confirmed using genome-wide data (39, 87, 127, 137), demonstrates that GC-alleles have a higher probability of fixation than do AT-alleles (for a discussion of possible artifacts, see 35, 61). Thus, according to polymorphism data, the evolution of GC-content in mammalian genomes is not simply driven by patterns of mutation.

The Origin of Isochores: The gBGC Model

Given that neither selectionist nor mutational models provide satisfactory explanations for the origin of GC-rich isochores, we focused on a third possible hypothesis: biased gene conversion (54). This model originates from work by Brown & Jiricny (20), who noted that the GC-bias of the mismatch repair machinery might lead to a gene conversion bias favoring GC-alleles (gBGC). A few years later, Holmquist (64) and Eyre-Walker (40) proposed the hypothesis that gBGC could be responsible for the evolution of GC-rich isochores. One strong prediction of this model is that sequences subject to a high level of recombination should be GC-rich. And indeed, the analysis of human sequences demonstrated a genome-wide positive correlation between crossover rate and GC-content ($r^2 = 0.15$ at the Mb scale) (48, 80). The mouse *Fxy* gene provides a spectacular example of this relationship. This gene was recently translocated into the pseudoautosomal region of the X chromosome, where the rate of crossover is extremely high. This translocation was followed by a strong acceleration in substitution rates (111) and a very strong increase of its GC-content at both coding and noncoding sites (e.g., a change in GC-content at third codon positions from 56% to 87%, in less than 3 million years) (52, 97). Nonallelic recombination also appears to be associated with increases

in GC-content. Indeed, in both mammals and birds, multigene families that are subject to frequent ectopic recombination (e.g., rRNAs, histones, HSP90, . . .) harbor an elevated GC-content (5, 49, 54, 81).

Recombination rates vary greatly among mammals (28), which implies that this parameter evolves very rapidly. For example, the average crossover rate is two times higher in humans than it is in mice (28). Conversely, the evolution of GC-content is generally a slow process (see below). Hence, to quantify correctly the relationship between recombination rate and the evolution of GC-content we must use estimates of recombination rates and substitution patterns measured on similar time scales. For this purpose, we can compare noncoding sequences (introns and intergenic regions) from closely related species to measure neutral base substitution rates during recent evolutionary times. From these substitution rates we can compute the stationary GC-content, i.e., the GC-content that sequences would reach at equilibrium if patterns of substitution remained constant over time (hereafter noted GC*). GC* can be considered as a summary statistic of the average substitution matrix since the divergence of the species being compared, and provides information about the recent trend of evolution of GC-content. By using this approach, we demonstrated that in primates, GC* is strongly correlated to the rate of crossover ($r^2 = 0.36$ at the Mb scale) (35, 95) (**Figure 4**). However, this is an underestimate of the true relationship between recombination and the evolution of GC-content, because (*a*) measures of crossover rates are noisy and (*b*) crossover rates are not a perfect estimate of the total recombination rate (crossover + noncrossover) (35). By using the distance to telomeres as an additional estimator of recombination rate, we showed that recombination explains at least 47% of the variance in GC* at the 1-Mb scale (35). The correlation between GC-content and recombination might partly be explained by a recombinogenic effect of GC-rich motifs (127). However, the fact that crossover rates are much more strongly correlated to GC* ($r^2 = 0.36$) than to the present

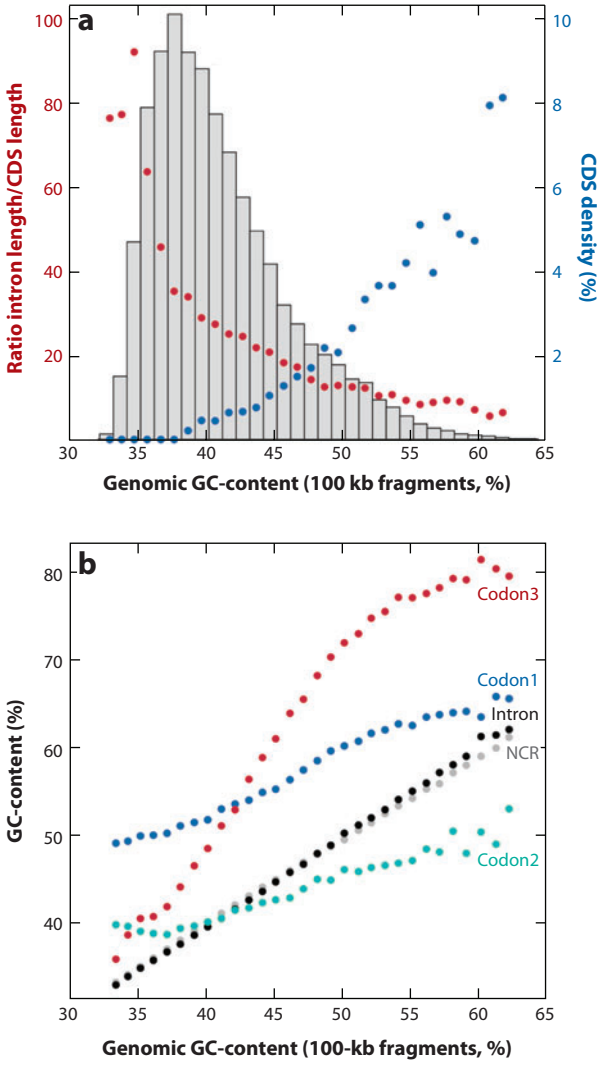

Figure 3

GC-content and gene density in 100-kb segments across the human genome. Coding density is much higher in GC-rich than in GC-poor regions, not only because intergenic regions are shorter but also because genes are more compact (have a lower intron/CDS length ratio). Variations in GC-content affect not only noncoding regions (introns, intergenic regions) but also coding exons, even at first and second codon positions (which indicates that the amino-acid composition of proteins depends partly on the GC-content of the region where the gene is located). (*a*) Histogram: frequency distribution of GC-content. Left axis: median ratio of intron length to coding DNA sequence (CDS) length. Right axis: median CDS density. (*b*) Median GC-content in different genomic compartments: introns of protein-coding genes, other noncoding regions (NCR), first, second, and third codon positions.

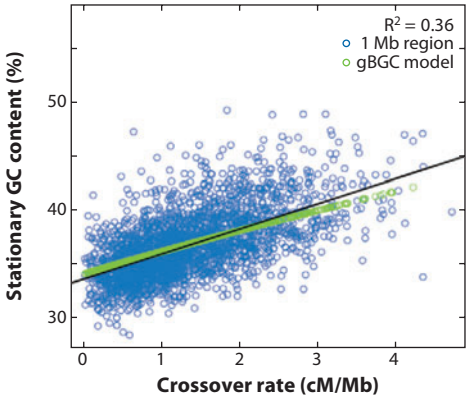

Figure 4

Correlation between the stationary GC-content and the crossover rate in human autosomes. Each dot corresponds to a 1-Mb-long genomic region. Green dots correspond to the predictions of the gBGC model. Data from Reference (35).

GC-content ($r^2 = 0.15$) clearly demonstrates that meiotic recombination has a major influence on the evolution of base composition, in agreement with the gBGC model (35, 95).

Another strong prediction of the gBGC model is that the fixation bias in favor of GC-rich alleles should be maximal in regions of high recombination. Again, this prediction fits with the observations that the derived allele frequency of GC-alleles is maximal within recombination hotspots and minimal in cold spots, whereas AT-alleles show the opposite pattern (89, 126). This pattern cannot be explained either by a mutagenic effect of recombination, or by artifacts in the polarization of polymorphism (35).

None of these observations fit with the predictions of selectionist models of isochore evolution (35, 126). Because of Hill-Robertson interference, the efficiency of selection is expected to be higher in regions of high crossover rate. However, this effect is very weak and unlikely to explain the strong correlation observed between GC* and crossover rate (35). Moreover, Hill-Robertson interference depends on the total rate of crossover in populations, in both females and males. Thus it predicts similar correlations between GC* and the male and female crossover rates. The finding that GC* is much more strongly correlated to male than to female crossover rates (see below) therefore rules out selective models (35).

Ectopic recombination: recombination between nonallelic homologous sequences

GC*: stationary GC-content

Theoretical Validation of the gBGC Model

The above observations are consistent with the predictions of the gBGC model. However, several authors have raised important objections to this model. First, Eyre-Walker (41) noted that 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. Second, Spencer et al. (127) raised doubt about the potential effect of gBGC on the human genome. In humans, recombination occurs predominantly in hotspots (typically 2 kb long) that cover about 3% of the genome (101). Moreover, the location of recombination hotspots is not conserved between human and chimpanzee, which indicates that hotspots have relatively short lifespans (118, 141). Finally, the fixation bias in favor of GC-alleles is relatively weak (127). Given this weakness and the spatial and temporal fluctuations in recombination rate, is it plausible that the gBGC process could affect the evolution of base composition in Mb-long genomic fragments?

This question was addressed by comparing the prediction of a simple model of genome evolution with gBGC to the pattern of substitution observed in the human lineage, since the divergence from chimpanzee (35). In this model, the pattern of mutation was assumed to be constant across the genome, with gBGC affecting only recombination hotspots. The model assumed that at the Mb scale, the average recombination rate remained constant during the period of time considered (i.e., hotspots have a short lifespan, but the hotspot density is maintained). The predictions of this simple model, with realistic assumptions regarding distribution of hotspots, fixation bias, and population size, fit remarkably with the observations (35) (**Figure 4**). Notably, this model explains the

relationships between recombination rates and substitution rates (35). It predicts that in an average human recombination hotspot, GC* should be close to 100% and substitution rates should be about two times higher than the genomic average. Given their short lifespans, most recombination hotspots do not induce substitution hotspots and they die well before their GC-content reaches 100% [which responds to the objection raised by Eyre-Walker (41)]. However, the gBGC model predicts that recombination hotspots that are very active and have a relatively long lifespan can lead to very strong (and GC-biased) substitution hotspots (35), which fits well with the observed pattern of substitution hotspots in the human genome (32). Furthermore, given that gBGC occurs primarily in hotspots, this model explains why the base composition of GC-rich isochores is heterogeneous (105). In agreement with Spencer et al. (127), this model indicates that gBGC is at present not strong enough to maintain GC-rich isochores in the human genome, consistent with the observation that the GC-content of GC-rich isochores is indeed decreasing in primates (3, 4, 6, 34, 39, 86, 95, 138). The model, however, predicts that in species such as chicken, where chromosomal recombination rate (per Mb) is up to 20 times higher than in human, gBGC should be creating GC-rich isochores. Again, this prediction fits with the observations (136). The model also predicts that the strength of gBGC should vary with the effective population size (35), but there are not yet enough data to test this prediction.

The gBGC model can also explain the timing of isochore evolution (35). In the absence of gBGC, the evolution of base composition at neutral sites is a very slow process: the time necessary to halve the difference between the present GC-content and the equilibrium GC-content is about 470 million years. Conversely, when gBGC is effective, the evolution of GC-content can be much faster [e.g., 62 million years in a genomic region of high recombination rate (30 cM/Mb) in a species with large effective population size (N_e = 50,000)] (35). Thus, the gBGC model may explain the rapid evolution of GC-rich isochores in the amniote cenancestor (310 to 350 Mya) (65, 83), and the slow decay observed in many mammals (3, 4, 6, 34, 39, 86, 95, 138).

Male-Driven gBGC?

The analysis of substitution patterns in Alu repeats (139) and in substitution hotspots (32) suggested that the impact of recombination was sex-specific. Genome-wide analysis of neutral substitution patterns in human autosomes (1-Mb scale) confirmed that GC* is much more strongly correlated to male crossover rate (R^2 = 0.27) than to female crossover rate (R^2 = 0.15) (35).

This observation is reminiscent of so-called male-driven evolution in humans whereby mutation rates are higher in the male than in the female germline (91). However, this parallel is probably misleading. The male-mutation effect is likely due to the higher number of cell divisions in the male germline (91), whereas the impact of gBGC depends on the number of meiosis events (which is the same in the two lineages). Given that the rate of crossover in human autosomes is on average 65% higher in females than in males (80), a stronger effect of gBGC would be expected in females than in males.

Note, however, that the gBGC model predicts a positive correlation between GC* and the total recombination rate (crossover + noncrossovers), and the crossover/noncrossover ratio is known to vary across chromosomes. The weaker correlation between GC* and female crossover rate could therefore simply reflect a more widely varying noncrossover rate in females than in males. Alternatively, note that the timing of meiosis differs between the two sexes—whereas the production of spermatocytes is continuous in adult males, female meiosis remains suspended after crossovers have formed (in the fetal ovary), up to the time of fertilization, which is many years later (28)—and it is therefore possible that the repair of mismatches in heteroduplex DNA proceeds differently in males and in females, which could explain the male-specific gBGC.

GC Content, Transcription, and Gene Density

A striking feature of mammalian genomes is the much higher gene density in GC-rich than in GC-poor isochores (38, 86, 98) (**Figure 3**). This was speculated to be a direct consequence of gBGC: if transcription promotes recombination, then gene-dense regions should be more subject to gBGC and hence should have a higher GC-content (32). Experiments in mammalian cell cultures indicate that transcription can promote mitotic recombination (107). However, the correlation between crossover rate and gene density is extremely weak ($R^2 = 0.02$) (80), suggesting transcription has little, if any, impact on meiotic recombination. Moreover, the correlation between GC* and crossover rate is the same in introns and in intergenic regions (35). Hence, there is no evidence for a relationship between gBGC and transcription.

The strong increase in GC-content in the *Fxy* gene after its translocation to the highly recombining pseudoautosomal region was accompanied by large deletions within its introns (97). A propensity for GC-rich or highly recombining regions to be prone to deletions could explain the higher gene density in GC-rich isochores (97). There is evidence that the occurrence of indels (insertions or deletions) depends both on recombination rate and on local GC-content (perhaps because sequences of extreme base composition trigger replication slippage) (84, 90). There is, however, still no direct evidence that the rate of deletion exceeds the rate of insertion in GC-rich regions, and this model therefore remains speculative.

The Evolution of GC-Content: Impact of Chromosome Size

We have seen above that recombination appears to drive the evolution of GC content. But what drives the evolution of recombination? The determinants of recombination rates at the kilobase scale are not yet understood. Although some overrepresented sequence motifs have been identified in recombination hotspots (101, 102), it is not yet established whether mutations in these motifs affect the activity of recombination hotspots. That hotspot activity can be influenced by factors other than local sequence determinants (106) could explain why the location of hotspots is not conserved between human and chimpanzee, despite considerable sequence identity (118, 141). At the Mb scale, it is known that the average rate of recombination (which depends on the density of recombination hotspots and on their intensity) is strongly constrained by the size of chromosome arms. Indeed, in human and chicken the average rate of crossover is strongly correlated ($R^2 > 0.8$) to the size of chromosome arms (66, 72, 86) (**Figure 5**), because, as in many other organisms, at least one crossover per chromosome arm is required per meiosis (28, 110). In mammals, although the genome size is nearly constant, the number (and hence size) of chromosomes has evolved rapidly as a result of frequent chromosomal rearrangements (104). Across mammalian species, there is a strong correlation ($R^2 > 0.8$) between the number of chromosome arms (between 10 and 52) and the total crossover rate (28, 110). Thus, the evolution of the karyotype (by translocations, chromosomal fissions, fusions, etc.) appears to be a major determinant of the evolution of the crossover rate of chromosomes.

Several observations indicate that the evolution of GC-content in mammalian genomes is correlated to their karyotype evolution. First, there is a strong correlation between chromosome size and GC-content, in both human and chicken (**Figure 5**). GC-rich regions are prone to chromosomal breakage, which could partly explain this relationship (135). However, in human, chromosome size correlates much more strongly with GC* (the future GC-content) than with GC (**Figure 5*b*,*c***). This clearly demonstrates that chromosome size influences (indirectly) the evolution of GC-content, not the other way round. Human chromosome 2 results from a Robertsonian fusion event after the human/chimp split. Such fusions

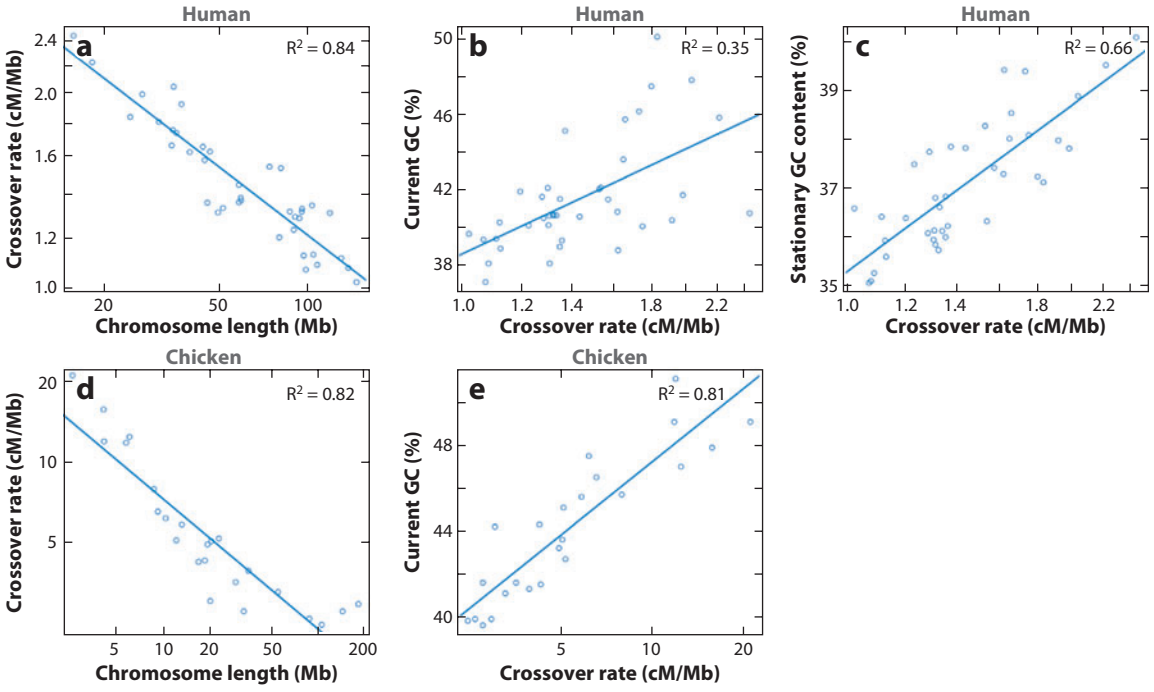

Figure 5

Correlations between chromosome length, crossover rate, and GC content in human and chicken autosomes. (*a-c*) Human chromosome arms. Average crossover rates in each chromosome arm were taken from HapMap (129). The stationary GC-content (GC), inferred from substitution rates in the human lineage, was computed with data from Reference (35). (*d-e*) Chicken chromosomes. Chicken crossover rates were taken from Reference (66). Chicken chromosome GGA16, GGA22, GGA23, GGA25, and GGA32 were excluded because of insufficient sequence coverage or insufficient genetic markers (66). Chromosome length and crossover rates are plotted in log scale. Regression lines and Pearson's correlation coefficients (R^2) are indicated.

increase the length of chromosomes and hence are expected to decrease recombination rates, and therefore to decrease the strength of gBGC. As predicted by this model, this fusion event is associated with a shift in substitution patterns, specifically in the human lineage (32). In fact, there is evidence that many chromosomal fusions occurred in the mammalian lineage (104), and it has been proposed that these fusions might be responsible for the erosion of GC-rich isochores observed in many mammalian genomes (6, 39). The most striking example of intensive fusions is the genome of the marsupial *Monodelphis domestica*, which consists of only 8 giant autosomes and one small X chromosome (96). As expected, these chromosomes have a very low crossover rate compared to eutherian mammals (except on the X) (28, 58), and the giant autosomes are AT-rich, whereas the small X is GC-rich (58). The opposite example is provided by the genome of the platypus (a monotreme), which comprises 52 chromosomes, and has a high GC-content (45.5%, compared to 40.9–41.8% in eutherians and 37.7% in *Monodelphis domestica*) (96, 132).

Another prediction of the gBGC model is that frequent chromosomal rearrangements should lead to a homogenization of GC-content (because the GC-content reflects the long-term average recombination rate). Thus the very high rate of chromosomal rearrangements in the rodent lineage (22, 55, 66) could explain why the distribution of GC-content is much more homogeneous in their genome compared to other eutherian mammals (the so-called minor-shift) (55, 100).

gBGC: A Universal Process?

So far, there is strong evidence for gBGC in yeast (15, 92) and in mammals (see above). Positive correlations between GC-content and recombination rates (estimated directly or inferred from chromosome size) have been reported in many other taxa: birds (66), turtle (83), nematode (93), *Drosophila* (94), paramecium (36), green alga (67), and plants (56) [note that there is one known exception to this rule (32a)]. Thus, the BGC process might be very widespread in eukaryotes. Most genomes, however, do not display the highly heterogeneous isochore organization observed in mammals. In fact, the evolution of isochores likely required a combination of different factors. First, BGC can be effective only if there is some heterozygosity. Thus, gBGC should have little or no effect in asexual or selfing species (56). Second, key gBGC parameters (recombination rate, conversion bias, i.e., the GC-bias in mismatch repair, and effective population size) must be large enough for gBGC to overcome genetic drift. Finally, gBGC should lead to variations of GC-content along sequences only if there is heterogeneity in recombination rates across chromosomes, and if the large-scale pattern of recombination remains relatively stable during evolution.

BIASED GENE CONVERSION AND NATURAL SELECTION

So far in this review, gBGC has been considered only in the context of neutral evolution. We have examined noncoding or silent positions, and, therefore, interpreted GC-biased departures from the neutral assumption as reflecting gBGC. Since the vast majority of sites in the mammalian genome are neutrally evolving, this rationale is appropriate for understanding the large-scale consequences of gBGC on genomic landscapes (isochores). On the other hand, functional, selected sites are obviously of primary interest to molecular and evolutionary biologists because they carry genetic information relevant to the development, function, and adaptation of organisms. In this section, we investigate the interaction between gBGC and natural selection. We highlight two problems posed by gBGC with respect to selection at the molecular level. The first is empirical: By generating nonneutral-like variation patterns, gBGC can confound selective inferences from genomic data. The second is fundamental: By influencing genetic evolution irrespective of the fitness of genotypes, gBGC can decrease the efficiency of natural selection in natural populations.

Biased Gene Conversion Mimics Natural Selection

By definition, gBGC results in the nonrandom transmission of alleles to the next generation. Consider a neutral position at which two alleles, A and G, are segregating in a population. In the absence of gBGC, each allele has a fixation probability equal to its current frequency (78). With gBGC, the G allele has an increased fixation probability (and the A allele a decreased probability), because heterozygotes produce, on average, more gametes carrying the G allele, thus tending to increase the population frequency of Gs generation after generation. Basic population genetic theory indicates that the population dynamics of the G allele under gBGC is identical to that of a positively selected allele (103): G has a population "advantage" over A. Of course, this advantage does not guarantee fixation, because stochastic variation of allele frequencies due to genetic drift can lead by chance to the fixation of the A allele, just as an advantageous mutation can be lost. This similarity between gBGC and selection has the consequence that gBGC-induced genomic variation patterns may be incorrectly interpreted as selection. We now review various examples of such confounding patterns in recent molecular evolutionary literature.

Within-species polymorphism. Within species, the main consequence of gBGC is a shift in allele frequency distribution: C and G alleles tend to segregate at higher frequency,

on average, than A and T. Such a pattern was reported in human from the analysis of single nucleotide polymorphism data (39, 41, 87, 127, 137). A similar pattern would be expected if G and C were selectively advantaged over A and T. The first reports of GC-biased polymorphism patterns in mammals were indeed interpreted this way (41).

The risk of confusion between selection and gBGC is especially high with codon usage patterns in Drosophila. In the *Drosophila* genome, as in many prokaryotic and eukaryotic genomes, synonymous codons—e.g. the two phenylalanine-encoding TTT and TTC—do not appear at equal frequencies: TTC is significantly more common than is TTT, especially in highly expressed genes (124). This reflects translational selection: TTC codons presumably are (slightly) more efficiently and/or accurately translated, as has been documented for decades in a number of prokaryotic and eukaryotic species (33, 123). This weak selective effect is apparent through the distribution of codon allele frequencies within populations: preferred codon alleles tend to segregate at higher frequency than do nonpreferred alleles (2). Kliman & Hey (79) showed that the magnitude of codon usage bias is higher, on average, for genes located in high-recombining regions of the *Drosophila* genome, and interpreted this pattern as reflecting Hill-Robertson interference (62): Multilocus natural selection is less effective in case of strong genetic linkage (109). Low-recombining regions, therefore, could accumulate slightly deleterious preferred → nonpreferred mutations at a rate higher than do high-recombining regions.

Codon usage in *Drosophila*, however, is peculiar in that for all 20 amino acids, the preferred codon is G- or C-ending (for a yet unexplained reason). So the preferred vs nonpreferred codon patterns could be at least partly explained by evolutionary processes influencing base composition, irrespective of translational selection. Marais et al. (93, 94) demonstrated that the GC-content of noncoding DNA was higher in high-recombining than in low-recombining regions of the *D. melanogaster* genome. They demonstrated that the positive relationship between codon usage and local recombination rate is explained by a neutral effect of recombination on base composition, not by Hill-Robertson interferences (and analyses in nematodes led to the same conclusion) (93, 94). Along the same lines, Galtier et al. (51) and Haddrill & Charlesworth (59) examined noncoding polymorphisms in *D. melanogaster* and *D. simulans*, and showed that the distribution of allele frequency is asymmetrical: In introns and intergenic DNA, G and C mutations reach higher frequency, on average, than do A and T mutations. These results indicate that the skewed allelic distribution of preferred vs nonpreferred codons (2) is partly due to nonselective causes: Preferred, GC-ending codons segregate at higher frequency not only because they are favorable, but also because they are promoted by a GC-biased evolutionary force affecting noncoding sequences as well, most probably gBGC. Analyzing codon usage patterns in three Brassicaceae species, Wright et al. (143) similarly concluded that variations in gBGC, not selection intensity, explained the shift in base composition in *Arabidopsis thaliana*.

HAR: human accelerated region

HACNS: human accelerated conserved noncoding sequence

Between-species divergence. Between species, gBGC can confound tests for selection because it affects the nucleotide substitution rate. When gBGC is active, AT → GC mutations have an elevated fixation probability, just like positively selected mutations. If an initially AT-rich sequence undergoes an episode of gBGC, a sudden burst of AT → GC substitutions and an increase in substitution rate should result (50). Accelerated evolution is typically interpreted as a signature of positive selection (144). gBGC episodes, therefore, could be falsely interpreted as adaptive events.

Recently, several research groups independently scanned the human genome in search of noncoding elements highly conserved across vertebrates, but highly divergent in human (14, 23, 77, 113, 116). Several such human-accelerated regions (HARs or HACNSs) were identified. These regions are most probably

functional, as suggested by their high level of conservation across nonhuman species. The accelerated evolution in the human lineage was interpreted as reflecting human-specific molecular adaptations. Consistently, the functional characterization of the most promising of these fast-evolving elements was discussed in adaptive terms (114, 117). A closer inspection of the variation pattern, however, revealed that (*a*) most of the nucleotide changes involved in these accelerated episodes are AT → GC, (*b*) human-accelerated sequences tend to be located in high-recombining regions of the human genome, and (*c*) the GC-biased pattern extends to flanking nonconserved regions (52). None of these features would be expected if positive selection were the driving force. Functional noncoding elements, for instance, are not GC-enriched, as compared to random noncoding sequences. We do not expect, therefore, that natural selection would generally favor AT → GC changes. These patterns, however, are expected under the gBGC model. This reanalysis suggests that gBGC, not adaptation, is the explanation for many of the human-specific accelerated events. The episodic nature of this process is consistent with the short life span of recombination hotspots in apes (118, 141).

Neglecting gBGC can lead to the spurious inference of selection when analyzing genomic data. We suggest that natural selection at the molecular level should be invoked only after both the neutral and the gBGC models (the extended null hypothesis of molecular evolution) have been rejected. This is particularly true in mammals and human, in which gBGC has influenced the evolution of a substantial fraction of the genes and genome. Distinguishing between gBGC and selection, however, is not always an easy task, as illustrated above. Three main features of gBGC can help distinguish it from selection. The first is that gBGC, unlike selection, only favors G and C alleles. Thus, when the observed nonneutral-like pattern is GC-biased, caution is advised in invoking selection. Second, gBGC applies to every position of a given region, irrespective of its functional status. When nonsynonymous, synonymous, and noncoding sequences are all affected, gBGC is the most likely explanation. Third, gBGC is mechanistically related to recombination. If the detected pattern is stronger in high-recombining regions, then again gBGC should be considered. None of these three criteria is entirely sufficient: Selection can occasionally be GC-biased (e.g., codon usage in *Drosophila*), affect noncoding sequences (e.g., HARs, HACNSs), and is theoretically related to recombination through the Hill-Robertson effect. The combination of all three criteria, however, is strongly suggestive of a gBGC, not selective effect. Most importantly, when the distinction between the two models' predictions are not clear-cut, gBGC remains a strong possibility and strong functional conclusions should not be drawn until it can be excluded.

Biased Gene Conversion Impedes Natural Selection

In the previous section, we highlighted an empirical problem: gBGC can confound our functional interpretations of genomic variation patterns by making neutral sequences look nonneutral. We now examine the consequences of gBGC on selected, functional sequences. How does gBGC interact with selection in natural populations? How does it affect the evolution of genes and regulatory regions? We first briefly outline theoretical expectations about the gBGC/selection interaction, and then we review existing data and relevant analyses.

Theoretical insights. Consider an AT → GC mutation with selection coefficient s (positive s for an advantageous mutation, negative s for a deleterious one). Assuming a gBGC process of strength b, this mutation would essentially behave like a mutation with selection coefficient $s+b$ (103). Similarly, a GC → AT selected mutation would behave under gBGC as if its selection coefficient, s, was diminished by b. So gBGC favors the fixation of advantageous AT → GC mutations, and the elimination of

Table 1 Interaction between GC-biased gene conversion and selection: a numerical example (effective population size $N_e = 10{,}000$; gBGC strength $N_e.b = 1$)

$N_e.s$	Type	Fixation probability no gBGC[a]	Fixation probability gBGC[a]	Frequency[b]	Substitution no gBGC[c]	Proportion no gBGC[d]	Substitution gBGC[c]	Proportion gBGC[d]
−10	AT → GC	$8.5\ 10^{-21}$	$4.2\ 10^{-19}$	0.125	0	**0.04**	0	**0.16**
−10	GC → AT	$8.5\ 10^{-21}$	$1.7\ 10^{-22}$	0.125	0		0	
−1	AT → GC	$3.7\ 10^{-6}$	$5.0\ 10^{-5}$	0.2	**1.5**		**20**	
−1	GC → AT	$3.7\ 10^{-6}$	$1.3\ 10^{-7}$	0.2	1.5		0.1	
0	AT → GC	$5.0\ 10^{-5}$	$2.0\ 10^{-4}$	0.15	15	0.42	61.1	0.51
0	GC → AT	$5.0\ 10^{-5}$	$3.7\ 10^{-6}$	0.15	15		1.1	
1	AT → GC	$2.0\ 10^{-4}$	$4.0\ 10^{-4}$	0.0225	9.2	0.54	18	0.33
1	GC → AT	$2.0\ 10^{-4}$	$5.0\ 10^{-5}$	0.0225	9.2		2.2	
10	AT → GC	$2.0\ 10^{-3}$	$2.2\ 10^{-3}$	0.0025	10		11	
10	GC → AT	$2.0\ 10^{-3}$	$1.8\ 10^{-3}$	0.0025	10		9	
Total				1.	71.2	1.	122.5	1.

[a] Fixation probability with/without gBGC.
[b] Assumed frequency distribution of mutations.
[c] Expected number of substitutions (for a gene of length 1000 sites undergoing total mutation rate 10^{-9} per site per generation during 10^7 generations) with or without gBGC.
[d] Proportion of deleterious, neutral and advantageous substitutions with or without gBGC.

deleterious GC → AT mutations, which is favorable to species adaptation. On the other hand, it decreases the fixation probability of advantageous GC → AT mutations, and increases that of deleterious AT → GC mutations, thus counteracting natural selection. On balance, the cost exceeds the benefit, as we now illustrate numerically with a theoretical example (**Table 1**) (see Reference 53 for detailed models and equations).

In this example, we analyze the fate of various classes of mutations in a population of effective size $N_e = 10{,}000$, without gBGC or with a relatively weak gBGC effect ($N_e.b = 1$). Mutations vary by their selection coefficient (from strongly deleterious to strongly advantageous) and AT↔GC character (**Table 1**). The third and fourth columns indicate that the fixation probabilities of neutral or weakly selected mutations are substantially affected by gBGC, whereas strongly selected mutations are not. Assuming an arbitrary distribution for the selective effect of mutations, in which deleterious and neutral mutations are much more common than advantageous ones (fifth column), the sixth and eighth columns reveal that the major effect of gBGC is a substantial increase in substitutions of weakly deleterious and neutral AT → GC mutations. As far as advantageous mutations are concerned, the modest increase in favorable AT → GC substitutions is essentially compensated by the decrease in favorable GC → AT ones. In this example, gBGC has two major consequences on the predicted evolution of selected sequences: an increase in total substitution rate, and an increased proportion of deleterious substitutions (from 4% to 16%). This simulation suggests that, by influencing the fate of mutations independently of their fitness effect, gBGC can counteract the effect of natural selection, and lead to the "undesired" fixation of deleterious mutations that would be eliminated in the absence of gBGC. In this theoretical example, the gBGC-induced substitution load contributes a substantial fraction of functional sequence evolution.

Empirical evidence. These theoretical arguments are only suggestive. Empirical analyses are required to investigate whether, and by how much, gBGC counteracts natural selection within populations, and negatively influences the evolution of functional sequences in nature. Some spectacular examples are available in the recent literature, as we now briefly review. The 3′ part of the mouse *Fxy* gene was recently translocated into the pseudoautosomal region, a recombination hotspot. Consequently, this gene was affected by a strong episode of gBGC, as shown by the very rapid increase in GC-content at synonymous and intronic positions (97) (see above). What about functional (nonsynonymous) sites? Consistent with theoretical predictions, the amino-acid substitution rate of this gene was strongly increased during the gBGC episode: 28 nonsynonymous substitutions, all from AT to GC, occurred in the mouse lineage posterior to the translocation (in less than 3 million years), whereas the total number of amino-acid differences between human and rat (diverged 80 Mya) is only four. In this example, strong gBGC has led to the fixation of slightly deleterious AT → GC mutations, which were counterselected prior to the gBGC episode (52). This example indicates that the strong correlation between the amino-acid composition of proteins and the GC-content of the genomic region where the gene is located (30) (**Figure 3**) might reflect, at least partly, the accumulation of weakly deleterious nonsynonymous substitutions driven by gBGC. Similarly, the GC-biased, human-accelerated functional noncoding elements [HARs and HACNSs (113, 116); see above] are consistent with the above theoretical predictions. These elements are generally under strong selective pressure, as revealed by their very low substitution rate in the absence of gBGC. When strong gBGC applies, moderately deleterious AT → GC changes reach substantial fixation probability, leading to a sudden increase in substitution rate (52). In the HAR1 noncoding RNA, these substitutions resulted in a profound rearrangement of the RNA structure (9). The functional consequences of deleterious evolution at locus HACNCS1 were experimentally characterized by Prabhakar et al. (117) in a mouse model. However, these authors interpreted their results in adaptive terms, a viewpoint with which we disagree (36a).

To approach the genomic impact of gBGC on coding sequence evolution, two research groups (9a, 53) recently searched for episodes of accelerated evolution in thousands of exonic sequences in human and apes. These studies showed that episodes of accelerated amino-acid evolution are more frequent in high-recombining regions of the genome, and are typically accompanied by an increase in GC-content, detectable at nonsynonymous sites, synonymous sites, and flanking intron sequences. The fact that the GC-bias and the accelerated evolution affected not only coding sites but also silent sites clearly shows that this pattern is not driven by positive selection. The nonsynonymous/synonymous rate ratio is generally increased during GC-biased episodes, consistent with a gBGC-induced relaxed selective pressure. Analyzing 46 nuclear genes in Triticeae (wheat and relatives), Haudry et al. (60) reported a higher nonsynonymous/synonymous ratio in outcrossing than in selfing species. The reverse pattern would be theoretically expected under selective models because of the reduced efficacy of purifying selection in selfers. Noting that the effect was stronger in GC-enriched and in high-recombining genes, these authors invoked gBGC as a potential explanation: gBGC is efficient in outcrossers, but not in selfing organisms, which are mainly homozygous. Accordingly, the gBGC-induced substitution load would explain the elevated rate of fixation of nonsynonymous mutations in outcrossers. Although these recent results warrant confirmation, they strongly suggest that gBGC, when active, can significantly (and negatively) affect the evolution of functional genomic elements, in human and in other species.

SUMMARY

The analysis of base composition and substitution patterns demonstrates that recombination has had a major influence on the evolution of GC-content in mammals and birds. This process, which is responsible for the evolution of peculiar genomic landscapes (the isochores) in amniotes, is not explainable by selection or a simple mutagenic effect. Patterns of polymorphism and substitutions in primates instead fit remarkably well with the hypothesis that the effect of recombination reflects GC-biased gene conversion (gBGC). This process, which has been directly demonstrated in yeast, is probably widespread in sexual eukaryotes.

In many of its effects, gBGC appears similar to selection. It is therefore essential to take this process into account in neutral models of sequence evolution to be able to detect selection within genomes. Moreover, the impact of gBGC on substitution patterns can be very strong, even in regions that are under selective pressure (coding sites or regulatory elements). Indeed, in some cases, gBGC overcomes purifying selection and leads to the fixation of deleterious AT → GC mutations (52, 53). Thus, recombination hotspots might constitute the Achilles' heels of our genome.

FUTURE ISSUES

GC-biased gene conversion is a recently discovered genetic process and its impact on genome evolution in humans and other species is only just starting to be characterized. Following are some of the issues opened by the above-reviewed literature that we think deserve to be addressed in the near future.

1. Molecular Mechanism

   One obvious question is to clarify the molecular mechanisms explaining gBGC. Which enzymes are involved? When during the recombination process does gBGC occur? Is it due to a recombination-initiation bias or a DNA repair bias? The occurrence of meiotic gBGC was experimentally demonstrated in yeast, but the situation is much less clear in human. Sperm typing in human did not yield any evidence for a generalized GC-bias, but this approach does not have the power at present to detect the moderate segregation biases sufficient to produce the evolutionary signature of gBGC. One especially intriguing feature is the much stronger correlation of gBGC to male than to female crossover rate in human, a pattern for which we currently have no satisfactory explanation.

2. Quantifying gBGC

   Molecular genetic studies would also help to quantify the strength of BGC and its genome-wide distribution. Notably, only a partial view of meiotic recombination in the human genome is available (we have data about the distribution of crossovers but not of noncrossovers), and the rate of mitotic gene conversion (which might also contribute to gBGC) is unknown. In yeast, the recently published estimates of recombination rate (crossover and noncrossover), gene conversion tract length, and GC-distortion rate (92) would yield a gBGC coefficient (b, see Box 2) of the order of 10^{-4} (genomic average), which suggests a strong effect of gBGC if the effective population size is 10^5 or higher. In human, an estimate of $4N_e.b = 1.3$ was obtained from population genetic data (127); this value applies to the top 20% high-recombining regions of the human genome. Levels of genetic resolution necessary to quantify gBGC are only starting to be achieved, especially given the strong spatio-temporal variations in recombination rate across genomes. The link between results obtained at the individual, population, and evolutionary levels cannot yet be made with any confidence.

3. gBGC, Codon Usage, and Gene Density
   The GC-content at synonymous codon positions is strongly correlated to the GC-content of flanking noncoding regions (30, 37) (**Figure 3**), which suggests that all classes of neutral sites are affected by gBGC. However, in GC-rich isochores, the GC-content is higher at synonymous codon positions than in introns (30, 37). Recombination hotspots are not concentrated within genes (101), and hence, there is a priori no reason why gBGC should affect exons more strongly than noncoding regions. Thus the gBGC model does not predict the observed difference in base composition between introns and synonymous sites. One possibility is that synonymous codon positions are under selective pressure (25, 41, 120). But why would genes in GC-poor isochores not be affected by this selective pressure? An alternative explanation is that the difference in base composition between synonymous sites and introns could be due to indels (including those resulting from the insertion of transposable elements), which are an important determinant of the evolution of GC-content in noncoding regions, but not in coding exons, where they are strongly counterselected (37). A better understanding of the determinants of indel rates is needed to investigate this issue and might also help us understand why gene density is strongly correlated to the regional GC-content.
4. Biased Gene Conversion and Comparative Genomic Patterns
   gBGC is a neutral force yielding nonneutral-like genomic patterns. We reviewed examples in which gBGC has confused selective inferences. Are these anecdotal instances, or is gBGC a general problem when detecting selective effects at the molecular level? We need to quantify the relative influences of gBGC and selection on genomic patterns. We know that gBGC partly explains codon usage patterns in *Drosophila* and *Caenorhabditis* (see above), but we do not know by how much, nor whether this effect applies to other species as well. Positive selection tests are also potentially affected. gBGC can influence the nonsynonymous/synonymous substitution rate ratio, and the distribution of allele frequencies within populations. We suggest that the various genome-scan analyses recently applied to detect targets of positive selection in human and other species be reexamined in the light of gBGC.
5. Functional Impact of Biased Gene Conversion
   By influencing the fixation probability of alleles independently of their fitness effect, gBGC can counteract natural selection. We are only starting to approach the evolutionary and functional consequences of the gBGC/selection interaction, both theoretically and empirically. We need to quantify the gBGC-induced substitution load, especially in human. What proportion of the amino-acid changes impacting our proteome is driven by gBGC, as compared to positive selection, or genetic drift? Are the examples of gBGC-driven deleterious evolution reported here only anecdotal, or are they the tip of a much bigger iceberg? It is tempting, for instance, to ask whether disease-causing deleterious mutations showing elevated prevalence in human are mainly AT → GC. The coanalysis of evolutionary and functional genomic data is needed to answer this question.
6. The Evolution of Biased Gene Conversion
   Direct or indirect evidence for gBGC has been reported in some animals, plants, fungi, and ciliates. As only a handful of genomes have been analyzed, the question of the taxonomic distribution of gBGC remains largely open. This is connected to the fundamental issue of the evolution of gBGC: if gBGC sometimes has deleterious consequences on

species adaptation, why did it evolve? Theoretical analyses indicate that BGC against the most common type of damage could reduce the mutational load, even if it occasionally favors deleterious mutations (8). Thus, Birdsell (15) suggested that gBGC evolved as an adaptation to a generalized AT-biased mutation process—a proposition that would need to be confirmed from comparative genome analyses. Note that it might not be necessary to invoke such second-order selective pressures to explain the evolution of gBGC: a selective pressure to limit the rate of somatic mutations should lead to the evolution of a GC-biased mismatch repair machinery, which, as a side effect, would generate gBGC when operating on mismatches in heteroduplex DNA. Finally, the long-term evolutionary consequences of gBGC are also of primary interest. gBGC is active when the recombination rate is high. If, as suggested above, gBGC has a negative effect on species adaptation by promoting the fixation of deleterious GC-alleles, then this could in turn influence the evolution of recombination, and of mating systems.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Bernard de Massy for very helpful comments on the manuscript. This work was supported by the Center National de la Recherche Scientifique, and by the Agence Nationale de la Recherche (ANR-08-GENM-036-01).

LITERATURE CITED

1. Aissani B, D'onofrio G, Mouchiroud D, Gardiner K, Gautier C, Bernardi G. 1991. The compositional properties of human genes. *J. Mol. Evol.* 32:493–503
2. Akashi H. 1997. Codon bias evolution in Drosophila. Population genetics of mutation-selection drift. *Gene* 205:269–78
3. Arndt PF, Hwa T, Petrov DA. 2005. Substantial regional variation in substitution rates in the human genome: importance of GC content, gene density, and telomere-specific effects. *J. Mol. Evol.* 60:748–63
4. Arndt PF, Petrov DA, Hwa T. 2003. Distinct changes of genomic biases in nucleotide substitution at the time of mammalian radiation. *Mol. Biol. Evol.* 20:1887–96
5. Backstrom N, Ceplitis H, Berlin S, Ellegren H. 2005. Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. *Mol. Biol. Evol.* 22:1992–99
6. Belle E, Duret L, Galtier N, Eyre-Walker A. 2004. The decline of isochores in mammals: an assessment of the GC content variation along the mammalian phylogeny. *J. Mol. Evol.* 58:653–60
7. Belle EM, Smith N, Eyre-Walker A. 2002. Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. *J. Mol. Evol.* 55:356–63
8. Bengtsson BO. 1986. Biased conversion as the primary function of recombination. *Genet. Res.* 47:77–80
9. Beniaminov A, Westhof E, Krol A. 2008. Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA* 14:1270–75

9a. Berglund J, Pollard KS, Webster MT. 2009. Hotspots of biased nucleotide substitutions in human genes. *PLoS Biol.* 7:e26

10. Bernardi G. 1993. The vertebrate genome: isochores and evolution. *Mol. Biol. Evol.* 10:186–204

Presents an analysis of protein-coding exons subject to episodes of accelerated evolution in mammals, which indicates that gBGC drives the fixation of deleterious mutations (see also 53).

11. Bernardi G. 2007. The neoselectionist theory of genome evolution. *Proc. Natl. Acad. Sci. USA* 104:8385–90
12. Bernardi G, Bernardi G. 1990. Compositional transitions in the nuclear genomes of cold-blooded vertebrates. *J. Mol. Evol.* 31:282–93
13. Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, et al. 1985. The mosaic genome of warm-blooded vertebrates. *Science* 228:953–58
14. Bird CP, Stranger BE, Liu M, Thomas DJ, Ingle CE, et al. 2007. Fast-evolving noncoding sequences in the human genome. *Genome Biol.* 8:R118
15. Birdsell JA. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol. Biol. Evol.* 19:1181–97
16. Boulton A, Myers RS, Redfield RJ. 1997. The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl. Acad. Sci. USA* 94:8058–63
17. Bradnam KR, Seoighe C, Sharp PM, Wolfe KH. 1999. G+C content variation along and among *Saccharomyces cerevisiae* chromosomes. *Mol. Biol. Evol.* 16:666–75
18. Brown TC, Jiricny J. 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* 50:945–50
19. Brown TC, Jiricny J. 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell* 54:705–11
20. Brown TC, Jiricny J. 1989. Repair of base-base mismatches in simian and human cells. *Genome* 31:578–83
21. Buard J, de Massy B. 2007. Playing hide and seek with mammalian meiotic crossover hotspots. *Trends Genet.* 23:301–9
22. Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, et al. 1999. The dynamics of chromosome evolution in birds and mammals. *Nature* 402:411–13
23. Bush EC, Lahn BT. 2008. A genome-wide screen for noncoding elements important in primate evolution. *BMC Evol. Biol.* 8:17
24. Caccio S, Perani P, Saccone S, Kadi F, Bernardi G. 1994. Single-copy sequence homology among the GC-richest isochores of the genomes from warm-blooded vertebrates. *J. Mol. Evol.* 39:331–39
25. Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* 7:98–108
26. Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat. Rev. Genet.* 8:762–75
27. Coop G, Myers SR. 2007. Live hot, die young: transmission distortion in recombination hotspots. *PLoS Genet.* 3:e35
28. Coop G, Przeworski M. 2007. An evolutionary view of human recombination. *Nat. Rev. Genet.* 8:23–34
29. Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. 2008. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* 319:1395–98
30. D'onofrio G, Mouchiroud D, Aissani B, Gautier C, Bernardi G. 1991. Correlation between the compositional properties of human genes, codon usage, and amino-acid composition of proteins. *J. Mol. Evol.* 32:504–10
31. de Massy B. 2003. Distribution of meiotic recombination sites. *Trends Genet.* 19:514–22
31a. Diaz RL, Alcid AD, Berger JM, Keeney S. 2002. Identification of residues in yeast Spo11p critical for meiotic DNA double-strand break formation. *Mol. Cell Biol.* 22:1106–15
32. Dreszer TR, Wall GD, Haussler D, Pollard KS. 2007. Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. *Genome Res.* 17:1420–30
32a. Drouaud J, Camilleri C, Bourguignon PY, Canaguier A, Berard A, et al. 2006. Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination "hot spots." *Genome Res.* 16:106–14
33. Duret L. 2002. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* 12:640–49
34. Duret L. 2006. The GC content of primates and rodents genomes is not at equilibrium: a reply to Antezana. *J. Mol. Evol.* 62:803–6
35. Duret L, Arndt PF. 2008. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* 4:e1000071

Many substitution hotspots in the human genome are caused by gBGC, notably associated to male recombination.

Demonstrates that the relationship between recombination and patterns of substitution in the human genome fits remarkably with the theoretical predictions of the gBGC model.

36. Duret L, Cohen J, Jubin C, Dessen P, Gout JF, et al. 2008. Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline. *Genome Res.* 18:585–96
36a. Duret L, Galtier N. 2009. Comment on "Human-specific gain of function in a developmental enhancer." *Science* 323:714
37. Duret L, Hurst LD. 2001. The elevated GC content at exonic third sites is not evidence against neutralist models of isochore evolution. *Mol. Biol. Evol.* 18:757–62
38. Duret L, Mouchiroud D, Gautier C. 1995. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. *J. Mol. Evol.* 40:308–17
39. Duret L, Semon M, Piganeau G, Mouchiroud D, Galtier N. 2002. Vanishing GC-rich isochores in mammalian genomes. *Genetics* 162:1837–47
40. Eyre-Walker A. 1993. Recombination and mammalian genome evolution. *Proc. R. Soc. London Ser. B* 252:237–43
41. 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* 152:675–83
42. Eyre-Walker A, Hurst LD. 2001. The evolution of isochores. *Nat. Rev. Genet.* 2:549–55
43. Federico C, Saccone S, Bernardi G. 1998. The gene-richest bands of human chromosomes replicate at the onset of the S-phase. *Cytogenet. Cell Genet.* 80:83–88
44. Filipski J. 1990. Evolution of DNA sequence contributions of mutational bias and selection to the origin of chromosomal compartments. *Adv. Mutagen. Res.* 2:1–54
45. Filipski J, Thiery JP, Bernardi G. 1973. An analysis of the bovine genome by Cs2SO4-Ag density gradient centrifugation. *J. Mol. Biol.* 80:177–97
46. Francino MP, Ochman H. 1999. Isochores result from mutation not selection. *Nature* 400:30–31
46a. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–61
47. Fryxell KJ, Zuckerkandl E. 2000. Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol. Biol. Evol.* 17:1371–83
48. Fullerton SM, Bernardo Carvalho A, Clark AG. 2001. Local rates of recombination are positively correlated with GC content in the human genome. *Mol. Biol. Evol.* 18:1139–42
49. Galtier N. 2003. Gene conversion drives GC content evolution in mammalian histones. *Trends Genet.* 19:65–68
50. Galtier N. 2004. Recombination, GC-content and the human pseudoautosomal boundary paradox. *Trends Genet.* 20:347–49
51. Galtier N, Bazin E, Bierne N. 2006. GC-biased segregation of noncoding polymorphisms in Drosophila. *Genetics* 172:221–28
52. Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* 23:273–77
53. Galtier N, Duret L, Glemin S, Ranwez S. 2009. GC-biased gene conversion promotes the fixation of deleterious amino-acid changes in primates. *Trends Genet.* 25:1–5
54. Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159:907–11
55. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, et al. 2004. Genome sequence of the brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521
56. Glemin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proc. Biol. Sci.* 273:3011–19
57. Deleted in proof
58. Goodstadt L, Heger A, Webber C, Ponting CP. 2007. An analysis of the gene complement of a marsupial, *Monodelphis domestica*: evolution of lineage-specific genes and giant chromosomes. *Genome Res.* 17:969–81
59. Haddrill PR, Charlesworth B. 2008. Non-neutral processes drive the nucleotide composition of noncoding sequences in Drosophila. *Biol. Lett.* 4:438–41
60. Haudry A, Cenci A, Guilhaumon C, Paux E, Poirier S, et al. 2008. Mating system and recombination affect molecular evolution in four Triticeae species. *Genet. Res.* 90:97–109

First evidence that at neutral sites, GC-alleles have a higher probability of fixation than AT-alleles.

The analysis of the evolution of the Fxy gene in mice and HARs in primates illustrates that gBGC can lead to the fixation of deleterious mutations and confound classical selection tests.

Presents a theoretical investigation of the interference between selection and gBGC and an analysis of protein-coding exons subject to episodes of accelerated evolution in mammals, which indicates that gBGC drives the fixation of deleterious mutations (see also 9a).

61. Hernandez RD, Williamson SH, Zhu L, Bustamante CD. 2007. Context-dependent mutation rates may cause spurious signatures of a fixation bias favoring higher GC-content in humans. *Mol. Biol. Evol.* 24:2196–202
62. Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* 8:269–94
63. Holloway K, Lawson VE, Jeffreys AJ. 2006. Allelic recombination and de novo deletions in sperm in the human beta-globin gene region. *Hum. Mol. Genet.* 15:1099–111
64. Holmquist GP. 1992. Chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.* 51:17–37
65. Hughes S, Zelus D, Mouchiroud D. 1999. Warm-blooded isochore structure in Nile crocodile and turtle. *Mol. Biol. Evol.* 16:1521–57
66. Int. Chicken Genome Sequencing Consort. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716
67. Jancek S, Gourbiere S, Moreau H, Piganeau G. 2008. Clues about the genetic basis of adaptation emerge from comparing the proteomes of two *Ostreococcus* ecotypes (Chlorophyta, Prasinophyceae). *Mol. Biol. Evol.* 25:2293–300
68. Jeffreys AJ, May CA. 2004. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat. Genet.* 36:151–56
69. Jeffreys AJ, Neumann R. 2002. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat. Genet.* 31:267–71
70. Jeffreys AJ, Neumann R. 2005. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.* 14:2277–87
71. Jonkman MF, Scheffer H, Stulp R, Pas HH, Nijenhuis M, et al. 1997. Revertant mosaicism in epidermolysis bullosa caused by mitotic gene conversion. *Cell* 88:543–51
72. Kaback DB. 1996. Chromosome-size dependent control of meiotic recombination in humans. *Nat. Genet.* 13:20–21
73. Kadi F, Mouchiroud D, Sabeur G, Bernardi G. 1993. The compositional patterns of the avian genomes and their evolutionary implications. *J. Mol. Evol.* 37:544–51
74. Kadyk LC, Hartwell LH. 1992. Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. *Genetics* 132:387–402
75. Kanaar R, Hoeijmakers JH, van Gent DC. 1998. Molecular mechanisms of DNA double strand break repair. *Trends Cell Biol.* 8:483–89
76. Khelifi A, Meunier J, Duret L, Mouchiroud D. 2006. GC content evolution of the human and mouse genomes: insights from the study of processed pseudogenes in regions of different recombination rates. *J. Mol. Evol.* 62:745–52
77. Kim SY, Pritchard JK. 2007. Adaptive evolution of conserved noncoding elements in mammals. *PLoS Genet.* 3:1572–86
78. Kimura M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–19
79. Kliman RM, Hey J. 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol. Biol. Evol.* 10:1239–58
80. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, et al. 2002. A high-resolution recombination map of the human genome. *Nat. Genet.* 31:241–47
81. Kudla G, Helwak A, Lipinski L. 2004. Gene conversion and GC-content evolution in mammalian Hsp70. *Mol. Biol. Evol.* 21:1438–44
82. Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol.* 4:e180
83. Kuraku S, Ishijima J, Nishida-Umehara C, Agata K, Kuratani S, Matsuda Y. 2006. cDNA-based gene mapping and GC3 profiling in the soft-shelled turtle suggest a chromosomal size-dependent GC bias shared by sauropsids. *Chromosome Res.* 14:187–202
84. Kvikstad EM, Tyekucheva S, Chiaromonte F, Makova KD. 2007. A macaque's-eye view of human insertions and deletions: differences in mechanisms. *PLoS Comput. Biol.* 3:1772–82
85. Lamb BC. 1984. The properties of meiotic gene conversion important in its effects on evolution. *Heredity* 53:113–38

86. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
87. Lercher MJ, Smith NG, Eyre-Walker A, Hurst LD. 2002. The evolution of isochores: evidence from SNP frequency distributions. *Genetics* 162:1805–10
88. Lercher MJ, Urrutia AO, Pavlicek A, Hurst LD. 2003. A unification of mosaic structures in the human genome. *Hum. Mol. Genet.* 12:2411–15
89. Li MK, Gu L, Chen SS, Dai JQ, Tao SH. 2008. Evolution of the isochore structure in the scale of chromosome: insight from the mutation bias and fixation bias. *J. Evol. Biol.* 21:173–82
90. Lunter G, Ponting CP, Hein J. 2006. Genome-wide identification of human functional DNA using a neutral indel model. *PLoS Comput. Biol.* 2:e5
91. Makova KD, Li WH. 2002. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* 416:624–26
92. Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454:479–85
93. Marais G, Mouchiroud D, Duret L. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl. Acad. Sci. USA* 98:5688–92
94. Marais G, Mouchiroud D, Duret L. 2003. Neutral effect of recombination on base composition in Drosophila. *Genet. Res.* 81:79–89
95. Meunier J, Duret L. 2004. Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.* 21:984–90
96. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167–77
97. Montoya-Burgos JI, Boursot P, Galtier N. 2003. Recombination explains isochores in mammalian genomes. *Trends Genet.* 19:128–30
98. Mouchiroud D, D'onofrio G, Aissani B, Macaya G, Gautier C, Bernardi G. 1991. The distribution of genes in the human genome. *Gene* 100:181–87
99. Mouchiroud D, Fichant G, Bernardi G. 1987. Compositional compartmentalization and gene composition in the genome of vertebrates. *J. Mol. Evol.* 26:198–204
100. Mouchiroud D, Gautier C, Bernardi G. 1988. The compositional distribution of coding sequences and DNA molecules in humans and murids. *J. Mol. Evol.* 27:311–20
101. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310:321–24
102. Myers S, Freeman C, Auton A, Donnelly P, McVean G. 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.* 40:1124–29
103. Nagylaki T. 1983. Evolution of a finite population under gene conversion. *Proc. Natl. Acad. Sci. USA* 80:6278–681
104. Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res.* 17:1254–65
105. Nekrutenko A, Li WH. 2000. Assessment of compositional heterogeneity within and between eukaryotic genomes. *Genome. Res.* 10:1986–95
106. Neumann R, Jeffreys AJ. 2006. Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. *Hum. Mol. Genet.* 15:1401–11
107. Nickoloff JA. 1992. Transcription enhances intrachromosomal homologous recombination in mammalian cells. *Mol. Cell. Biol.* 12:5311–18
108. Olsen AK, Bjortuft H, Wiger R, Holme J, Seeberg E, et al. 2001. Highly efficient base excision repair (BER) in human and rat male germ cells. *Nucleic Acids Res.* 29:1781–90
109. Otto SP, Barton NH. 1997. The evolution of recombination: removing the limits to natural selection. *Genetics* 147:879–906
110. Pardo-manuel De Villena F, Sapienza C. 2001. Recombination is proportional to the number of chromosome arms in mammals. *Mamm. Genome.* 12:318–22
111. Perry J, Ashworth A. 1999. Evolutionary rate of a gene affected by chromosomal position. *Curr. Biol.* 9:987–89

The genome-wide analysis of meiotic recombination products uncovers the existence of gBGC in yeast.

The first demonstration of the strong impact of recombination on the evolution of GC-content in primates.

112. Piganeau G, Westrelin R, Tourancheau B, Gautier C. 2001. Multiplicative versus additive selection in relation to genome evolution: a simulation study. *Genet. Res.* 78:171–75
113. Pollard KS, Salama SR, King B, Kern AD, Dreszer T, et al. 2006. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* 2:e168
114. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, et al. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443:167–72
115. Pop R, Zaragoza MV, Gaudette M, Dohrmann U, Scherer G. 2005. A homozygous nonsense mutation in SOX9 in the dominant disorder campomelic dysplasia: a case of mitotic gene conversion. *Hum. Genet.* 117:43–53
116. Prabhakar S, Noonan JP, Paabo S, Rubin EM. 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science* 314:786
117. Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, et al. 2008. Human-specific gain of function in a developmental enhancer. *Science* 321:1346–50
118. Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, et al. 2005. Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.* 37:429–34
119. Ream RA, Johns GC, Somero GN. 2003. Base compositions of genes encoding alpha-actin and lactate dehydrogenase-A from differently adapted vertebrates show no temperature-adaptive variation in G + C content. *Mol. Biol. Evol.* 20:105–10
120. Resch AM, Carmel L, Marino-Ramirez L, Ogurtsov AY, Shabalina SA, et al. 2007. Widespread positive selection in synonymous sites of mammalian genes. *Mol. Biol. Evol.* 24:1821–31
121. Semon M, Lobry JR, Duret L. 2006. No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Mol. Biol. Evol.* 23:523–29
122. Semon M, Mouchiroud D, Duret L. 2005. Relationship between gene expression and GC-content in mammals: statistical significance and biological relevance. *Hum. Mol. Genet.* 14:421–27
123. Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. 1995. DNA sequence evolution: the sounds of silence. *Philos. Trans. R. Soc. London Ser. B* 349:241–47
124. Shields DC, Sharp PM, Higgins DG, Wright F. 1988. "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* 5:704–16
125. Smit AF. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* 9:657–63
126. Spencer CC. 2006. Human polymorphism around recombination hotspots. *Biochem. Soc. Trans* 34:535–36
127. Spencer CC, Deloukas P, Hunt S, Mullikin J, Myers S, et al. 2006. The influence of recombination on human genetic diversity. *PLoS Genet.* 2:e148
128. Surtees JA, Argueso JL, Alani E. 2004. Mismatch repair proteins: key regulators of genetic recombination. *Cytogenet. Genome Res.* 107:146–59
129. The Int. HapMap Consort. 2005. A haplotype map of the human genome. *Nature* 437:1299–320
130. Vinogradov AE. 2003. Isochores and tissue-specificity. *Nucleic Acids Res.* 31:5212–20
131. Vinogradov AE. 2005. Noncoding DNA, isochores and gene expression: nucleosome formation potential. *Nucleic Acids Res.* 33:559–63
132. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175–83
133. Watanabe Y, Fujiyama A, Ichiba Y, Hattori M, Yada T, et al. 2002. Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. *Hum. Mol. Genet.* 11:13–21
134. Webb AJ, Berg IL, Jeffreys A. 2008. Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. *Proc. Natl. Acad. Sci. USA* 105:10471–76
135. Webber C, Ponting CP. 2005. Hotspots of mutation and breakage in dog and human chromosomes. *Genome Res.* 15:1787–97
136. Webster MT, Axelsson E, Ellegren H. 2006. Strong regional biases in nucleotide substitution in the chicken genome. *Mol. Biol. Evol.* 23:1203–16
137. Webster MT, Smith NG. 2004. Fixation biases affecting human SNPs. *Trends Genet.* 20:122–26
138. Webster MT, Smith NG, Ellegren H. 2003. Compositional evolution of noncoding DNA in the human and chimpanzee genomes. *Mol. Biol. Evol.* 20:278–86

Presents the first evidence of the influence of recombination hotspots on genetic variation and the fate of individual mutations.

Human recombination hotspots, analyzed by sperm-typing, frequently show evidence of BGC and of polymorphism in crossover activity.

139. Webster MT, Smith NG, Hultin-Rosenberg L, Arndt PF, Ellegren H. 2005. Male-driven biased gene conversion governs the evolution of base composition in human alu repeats. *Mol. Biol. Evol.* 22:1468–74

Presents the first evidence of male-driven gBGC in the human genome.

140. Whitby MC. 2005. Making crossovers during meiosis. *Biochem. Soc. Trans* 33:1451–55
141. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, et al. 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308:107–11
142. Wolfe KH, Sharp PM, Li WH. 1989. Mutation rates differ among regions of the mammalian genome. *Nature* 337:283–85
143. Wright SI, Iorgovan G, Misra S, Mokhtari M. 2007. Neutral evolution of synonymous base composition in the Brassicaceae. *J. Mol. Evol.* 64:136–41
144. Yang Z. 2002. Inference of selection from multiple species alignments. *Curr. Opin. Genet. Dev.* 12:688–94

Contents

Annual Review of Genomics and Human Genetics

Volume 10, 2009

Chromosomes in Leukemia and Beyond: From Irrelevant to Central Players
Janet D. Rowley 1

Unraveling a Multifactorial Late-Onset Disease: From Genetic Susceptibility to Disease Mechanisms for Age-Related Macular Degeneration
Anand Swaroop, Emily Y. Chew, Catherine Bowes Rickman, and Gonçalo R. Abecasis ...19

Syndromes of Telomere Shortening
Mary Armanios 45

Chronic Pancreatitis: Genetics and Pathogenesis
Jian-Min Chen and Claude Férec 63

The Genetics of Crohn's Disease
Johan Van Limbergen, David C. Wilson, and Jack Satsangi 89

Genotyping Technologies for Genetic Research
Jiannis Ragoussis 117

Applications of New Sequencing Technologies for Transcriptome Analysis
Olena Morozova, Martin Hirst, and Marco A. Marra 135

The Posttranslational Processing of Prelamin A and Disease
Brandon S.J. Davies, Loren G. Fong, Shao H. Yang, Catherine Coffinier, and Stephen G. Young 153

Genetic Testing in Israel: An Overview
Guy Rosner, Serena Rosner, and Avi Orr-Urtreger 175

Stewardship of Human Biospecimens, DNA, Genotype, and Clinical Data in the GWAS Era
Stephen J. O'Brien 193

Schistosoma Genetics: New Perspectives on Schistosome Biology and Host-Parasite Interaction
Ze-Guang Han, Paul J. Brindley, Sheng-Yue Wang, and Zhu Chen 211

Evolution of Genomic Imprinting: Insights from Marsupials and Monotremes
Marilyn B. Renfree, Timothy A. Hore, Geoffrey Shaw, Jennifer A. Marshall Graves, and Andrew J. Pask 241

Methods for Genomic Partitioning
Emily H. Turner, Sarah B. Ng, Deborah A. Nickerson, and Jay Shendure 263

Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes
Laurent Duret and Nicolas Galtier 285

Inherited Variation in Gene Expression
Daniel A. Skelly, James Ronald, and Joshua M. Akey 313

Genomic Analyses of Sex Chromosome Evolution
Melissa A. Wilson and Kateryna D. Makova 333

Sequencing Primate Genomes: What Have We Learned?
Tomas Marques-Bonet, Oliver A. Ryder, and Evan E. Eichler 355

Genotype Imputation
Yun Li, Cristen Willer, Serena Sanna, and Gonçalo Abecasis 387

Genetics of Athletic Performance
Elaine A. Ostrander, Heather J. Huson, and Gary K. Ostrander 407

Genetic Screening for Low-Penetrance Variants in Protein-Coding Diseases
Jill Waalen and Ernest Beutler 431

Copy Number Variation in Human Health, Disease, and Evolution
Feng Zhang, Wenli Gu, Matthew E. Hurles, and James R. Lupski 451

The Toxicogenomic Multiverse: Convergent Recruitment of Proteins Into Animal Venoms
Bryan G. Fry, Kim Roelants, Donald E. Champagne, Holger Scheib, Joel D.A. Tyndall, Glenn F. King, Timo J. Nevalainen, Janette A. Norman, Richard J. Lewis, Raymond S. Norton, Camila Renjifo, and Ricardo C. Rodríguez de la Vega 483

Genetics, Medicine, and the Plain People
Kevin A. Strauss and Erik G. Puffenberger 513

Indexes

Cumulative Index of Contributing Authors, Volumes 1–10 537

Cumulative Index of Chapter Titles, Volumes 1–10 541

Errata

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at http://genom.annualreviews.org/