Surprising Fitness Consequences of GC-Biased Gene Conversion: I. Mutation Load and Inbreeding Depression

Sylvain Glémin¹

Institut des Sciences de l'Evolution, Université Montpellier II, Montpellier 34095, France

Manuscript received March 5, 2010

Accepted for publication April 19, 2010

ABSTRACT

GC-biased gene conversion (gBGC) is a recombination-associated process mimicking selection in favor of G and C alleles. It is increasingly recognized as a widespread force in shaping the genomic nucleotide landscape. In recombination hotspots, gBGC can lead to bursts of fixation of GC nucleotides and to accelerated nucleotide substitution rates. It was recently shown that these episodes of strong gBGC could give spurious signatures of adaptation and/or relaxed selection. There is also evidence that gBGC could drive the fixation of deleterious amino acid mutations in some primate genes. This raises the question of the potential fitness effects of gBGC. While gBGC has been metaphorically termed the “Achilles’ heel” of our genome, we do not know whether interference between gBGC and selection merely has practical consequences for the analysis of sequence data or whether it has broader fundamental implications for individuals and populations. I developed a population genetics model to predict the consequences of gBGC on the mutation load and inbreeding depression. I also used estimates available for humans to quantitatively evaluate the fitness impact of gBGC. Surprising features emerged from this model: (i) Contrary to classical mutation load models, gBGC generates a fixation load independent of population size and could contribute to a significant part of the load; (ii) gBGC can maintain recessive deleterious mutations for a long time at intermediate frequency, in a similar way to overdominance, and these mutations generate high inbreeding depression, even if they are slightly deleterious; (iii) since mating systems affect both the selection efficacy and gBGC intensity, gBGC challenges classical predictions concerning the interaction between mating systems and deleterious mutations, and gBGC could constitute an additional cost of outcrossing; and (iv) if mutations are biased toward A and T alleles, very low gBGC levels can reduce the load. A robust prediction is that the gBGC level minimizing the load depends only on the mutational bias and population size. These surprising results suggest that gBGC may have nonnegligible fitness consequences and could play a significant role in the evolution of genetic systems. They also shed light on the evolution of gBGC itself.

GC-BIASED gene conversion (gBGC) is increasingly recognized as a widespread force in shaping genome evolution. In different species, gene conversion occurring during double-strand break recombination repair is thought to be biased toward G and C alleles. In heterozygotes, GC alleles undergo a kind of molecular meiotic drive that mimics selection (reviewed in MARAIS 2003). This process can rapidly increase the GC content, especially around recombination hotspots (SPENCER *et al.* 2006), and, more broadly, can affect genome-wide nucleotide landscapes (DURET and GALTIER 2009a). For instance, it is thought to play a role in shaping isochore structure evolution in mammals (GALTIER *et al.* 2001; MEUNIER and DURET 2004; DURET *et al.* 2006) and birds (WEBSTER *et al.* 2006). Direct experimental evidence of gBGC mainly comes from studies in yeast (BIRDELL 2002; MANCERA

et al. 2008; but see MARSOLIER-KERGOAT and YERAMIAN 2009) and humans (BROWN and JIRICNY 1987). However, associations between recombination and the nucleotide landscape and frequency spectra biased toward GC alleles provide indirect evidence in very diverse organisms (Table 1).

The impact of gBGC on noncoding sequences and synonymous sites has been studied in depth, especially because of confounding effects with selection on codon usage (MARAIS *et al.* 2001). More recently, GALTIER and DURET (2007) pointed out that gBGC may also interfere with selection when affecting functional sequences. They argued that gBGC could leave spurious signatures of adaptive selection and proposed to extend the null hypothesis of molecular evolution. Indeed, gBGC can lead to a ratio of nonsynonymous (d_N) over synonymous (d_S) substitutions above one (BERGLUND *et al.* 2009; GALTIER *et al.* 2009), *i.e.*, a typical signature of positive selection (NIELSEN 2005). This hypothesis has been widely debated for human-accelerated regions (HARs). These regions are extremely conserved across mammals

¹Address for correspondence: Institut des Sciences de l'Evolution, UMR 5554, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France. E-mail: glemin@univ-montp2.fr

TABLE 1
Direct and indirect evidence of gBGC in several organisms

Organisms	Direct evidence	Indirect evidence	Achille's heel evidence	References
Yeast	Meiotic segregation bias Mitotic and mitotic heteromismatch correction bias	Correlation between GC and recombination		MANCERA <i>et al.</i> (2008) BIRDELL (2002)
Mammals	Mitotic heteromismatch correction bias	Correlation between GC*/GC and recombination Biased frequency spectrum toward GC alleles	GC bias associated with high d_N/d_S near recombination hotspot	BROWN and JIRICNY (1987) DURET and ARNDT (2008); MEUNIER and DURET (2004) GALTIER <i>et al.</i> (2001); SPENCER <i>et al.</i> (2006) BERGLUND <i>et al.</i> (2009); GALTIER <i>et al.</i> (2009)
Birds		Correlation between GC and recombination		INTERNATIONAL CHICKEN GENOME SEQUENCING CONSORTIUM (2004) KURAKU <i>et al.</i> (2006)
Turtles		Correlation between GC and chromosome size		
Drosophila		Correlation between GC and recombination Biased frequency spectrum toward GC alleles		MARAIS <i>et al.</i> (2003) GALTIER <i>et al.</i> (2006)
Nematodes		Correlation between GC and recombination		MARAIS <i>et al.</i> (2001)
Grasses		Correlation between GC and outcrossing/selfing Correlation between GC* and recombination and outcrossing/selfing	Outcrossing increases d_N/d_S for genes with high GC*	GLÉMIN <i>et al.</i> (2006) HAUDRY <i>et al.</i> (2008)
Green algae		Correlation between GC and recombination		JANCEK <i>et al.</i> (2008)
Paramecium		Correlation between GC and chromosome size		DURET <i>et al.</i> (2008)

but show evidence of accelerated evolution along the human lineage, which has been interpreted as evidence of positive selection (POLLARD *et al.* 2006a,b; PRABHAKAR *et al.* 2006, 2008). On the contrary, other authors argued that patterns observed in HARs, such as the AT → GC substitution bias, the absence of a selective sweep signature, or the propensity to occur within or close to recombination hotspots, are more likely explained by gBGC rather than positive selection (GALTIER and DURET 2007; BERGLUND *et al.* 2009; DURET and GALTIER 2009b; but see also POLLARD *et al.* 2006a who also suggested that gBGC might play a role in HARs evolution). It is thus crucial to take gBGC into account when interpreting genomic data.

Moreover, GALTIER and DURET (2007) initially suggested that gBGC hotspots could contribute to the fixation of slightly deleterious AT → GC mutations and could represent the Achilles' heel of our genome. This hypothesis was reinforced later in primates, with evidence of gBGC-driven fixation of deleterious mutations in proteins (GALTIER *et al.* 2009). A similar result

was also found in some grass species, whose genomes are also supposed to be affected by gBGC (Table 1; GLÉMIN *et al.* 2006). HAUDRY *et al.* (2008) compared two outcrossing and two selfing grass species and showed that GC-biased genes exhibit higher d_N/d_S ratio in outcrossing than in selfing lineages. The reverse pattern would be expected under pure selective models because of the reduced selection efficacy in selfers (CHARLESWORTH 1992; GLÉMIN 2007). This pattern is in agreement with a genomic Achilles' heel associated with outcrossing, while gBGC is inefficient in selfing species because they are mainly homozygous.

Twenty years ago, BENGTTSSON (1990) already pointed out that biased conversion can generally affect the mutation load. The mutation load is the reduction in the mean fitness of a population due to mutation accumulation, which could lead to population extinction if it is too high (LYNCH *et al.* 1995). At this time, Bengtsson concluded that "it is impossible to know if biased conversion plays a major role in determining the magnitude of the mutation load in organisms such as

ourselves, but the possibility must be considered and further investigated (BENGTTSSON 1990, p. 186).” Now, one can propose gBGC could be such a widespread biased conversion process. It thus appears timely to thoroughly investigate the fitness consequences of gBGC through its potential effects on the dynamics of deleterious mutations. The fitness consequences of gBGC were also pointed out as a major future issue to be addressed by DURET and GALTIER (2009a). In addition to the load, deleterious mutations have many other evolutionary consequences (for review see CHARLESWORTH and CHARLESWORTH 1998). They are thought to be the main determinant of inbreeding depression, *i.e.*, the reduction in fitness of inbred individuals compared to outbred ones. They also play a key role in the evolution of genetic systems (sexual reproduction and recombination, inbreeding avoidance mechanisms, ploidy cycles), of senescence, or in the degeneration of nonrecombining regions, such as Y chromosomes. So far, we know little, if anything, about how gBGC might affect these processes.

In his seminal work, BENGTTSSON (1990) did not address several important points. First, he did not include genetic drift in his model. Nearly neutral mutations, for which drift and selection are of similar intensities, are the most damaging ones because they can drift to fixation, unlike strongly deleterious mutations that are maintained at low frequency (CROW 1993; LANDE 1994, 1998). While gBGC intensities are rather weak (BIRDELL 2002; SPENCER *et al.* 2006), they could markedly affect the fate of nearly neutral mutations (see also GALTIER *et al.* 2009). Second, Bengtsson did not study the effect of gene conversion on inbreeding depression, while he showed that recessive mutations, mostly involved in inbreeding depression, are the most affected by gene conversion. Third, he did not envisage systematic GC bias with its opposite effects on A/T and G/C deleterious alleles. Fourth, while he noted that selfing affects both the efficacy of selection and that of conversion, he did not fully investigate the effect of mating systems. On one hand, selfing is efficient in purging strongly deleterious mutations causing inbreeding depression. However, since selfing is expected to increase drift, weakly deleterious mutations can fix in selfing species, contributing to the so-called “drift load” (CHARLESWORTH 1992; GLÉMIN 2007). Self-fertilizing populations are thus expected to exhibit low inbreeding depression and high drift load. On the other hand, gBGC, and thus its cost, vanishes as the selfing rate and homozygosity increase (MARAIS *et al.* 2004). gBGC could thus challenge classical views on mating systems and it was even speculated that gBGC could affect their evolution (HAUDRY *et al.* 2008).

Here I present a population genetics model that includes mutation, selection, drift, and gBGC, which extends previous studies (GUTZ and LESLIE 1976; LAMB and HELMI 1982; NAGYLAKI 1983a,b; BENGTTSSON 1990). I specifically examine how gBGC can affect inbreeding

depression and the mutation load. I also focus on the effect of mating system, which is especially interesting with regard to the interaction between biased conversion and selection. Finally, I discuss how these results could give insight into how gBGC evolved.

MODEL AND RESULTS

General formulation: Consider a single biallelic locus with weak ($W = A$ or T) and strong ($S = G$ or C) alleles. The life cycle is as follows: N diploid adults produce gametes after mutation and conversion events. Heterozygote individuals produce S alleles with probability $1/2(1 + b)$ and W alleles with probability $1/2(1 - b)$, where b is the gBGC coefficient (disparity coefficient *sensu* NAGYLAKI 1983a). Fertilization occurs and zygotes experience selection to give adults, which experience regulation to N individuals. If the S (resp. W) allele is deleterious, the relative fitnesses of WW (resp. SS), WS , and SS (resp. WW) genotypes are 1 , $1 - hs$, and $1 - s$, respectively, where s is the selection coefficient and h the dominance coefficient. The W allele mutates at rate u to the S allele. The reverse mutation occurs at rate $v = \lambda u$. λ is the mutation bias from S to W alleles, which ranges from 2 to 4.5 in many different organisms (LYNCH 2007). I assume that u and v are much smaller than the selection and conversion coefficients ($u, v \ll s, hs$, and b). For simplicity, I also assume that the gBGC intensity is constant. However, strong gBGC events are thought to be associated with short-lived recombination hotspots, at least in humans (MCVEAN *et al.* 2004; MYERS *et al.* 2005). I thus implicitly assume that gBGC/selection dynamics are shorter than the recombination hotspot lifespan. The validity of this assumption is discussed below. All parameters are summarized in Table 2.

Deterministic equilibrium—three selection regimes: Analysis of the deterministic case ($N = \infty$) highlights the different behaviors of the model. BENGTTSSON (1990) gave the full equation of the deterministic change in allele frequency. Under the assumption of weak selection and weak gBGC ($s \ll 1$ and $b \ll 1$), changes due to mutation, selection, and gBGC can be computed separately. Note that a full treatment of the system, without approximations, is given by A. POPA, S. GLEMIN, E. POPA, D. MOUCHIROUD, and C. GAUTIER (unpublished data). The deterministic change in frequency of the deleterious allele (x) can thus be approximated by

$$\Delta x = \Delta x_{\text{mutation}} + \Delta x_{\text{conversion}} + \Delta x_{\text{selection}} \quad (1)$$

with

$$\Delta x_{\text{mutation}} = \begin{cases} u(1 - x) - \lambda ux & \text{if } S \text{ is deleterious} \\ \lambda u(1 - x) - ux & \text{if } W \text{ is deleterious} \end{cases}$$

$$\Delta x_{\text{conversion}} = \begin{cases} b(1 - F)x(1 - x) & \text{if } S \text{ is deleterious} \\ -b(1 - F)x(1 - x) & \text{if } W \text{ is deleterious} \end{cases}$$

and

TABLE 2
Notations

N : Actual population size equals effective population size in panmictic populations
F : Wright's fixation index
α : Scaling factor summarizing the reduction in effective population size due to hitchhiking and bottleneck effects
u : Mutation rate from AT toward GC
v : Mutation rate from GC toward AT
$\lambda = v/u$: Mutation bias toward AT
h : Dominance coefficient of the deleterious allele
s : Selection coefficient against the deleterious allele
b : Intensity of biased gene conversion
q : Fraction of sites where GC is the advantageous allele ($q = \frac{1}{2}$ for nonsynonymous sites; q can be different from $\frac{1}{2}$ for synonymous sites/codon usage)
$s_{\max}$: Selection coefficient that maximizes the load
$b_{\min}^{\text{seg}}$: gBGC level that minimizes the segregating load
$b_{\min}^{\text{fix}}$: gBGC level that minimizes the fixation load
b_{lim} : gBGC level above which the load increases

$$\Delta x_{\text{selection}} = \frac{-sx(1-x)(F+h-hF+x(1-F)(1-2h))}{\bar{W}} \approx -sx(1-x)(F+h-hF+x(1-F)(1-2h)),$$

where $\bar{W} = 1 - 2x(1-x)(1-F)hs - (x^2 + Fx(1-x))s$ is the mean fitness of the population and F is Wright's fixation index, F_{IS} .

Note that in $\Delta x_{\text{selection}}$, the approximation $\bar{W} \approx 1$ holds either for $s \ll 1$ or $x \ll 1$. So it is still valid for lethal alleles because $x \ll 1$. For weak selection, the F coefficient can be used for neutral loci, that is, $F = \sigma/(2-\sigma)$, where σ is the selfing rate (CABALLERO and HILL 1992). For strong selection, F depends on selfing rate and selection parameters (GLÉMIN 2003). Given the approximation for $\Delta x_{\text{selection}}$, the effect of gBGC is equivalent to modifying the selection and dominance coefficients as follows:

$$\begin{aligned} s_{\text{equivalent}} &= s - 2\beta \\ h_{\text{equivalent}} &= \frac{\beta - hs}{2\beta - s} \end{aligned} \quad (2)$$

with

$$\beta = \begin{cases} b \frac{(1-F)}{(1+F)} & \text{if } S \text{ is deleterious} \\ -b \frac{(1-F)}{(1+F)} & \text{if } W \text{ is deleterious.} \end{cases}$$

Conventional population genetics results thus hold using this new parameterization. In panmixia ($F = 0$), gBGC is thus fully equivalent to modifying selection patterns. However, under inbreeding, the analogy is not complete because selection and dominance coefficients depend on F .

FIGURE 1.—Deterministic equilibrium frequency of the deleterious GC allele as a function of the conversion bias for different dominance levels: $h = 0.5$ (thick line), $h = 0.2$ (thin line), $h = 0.05$ (dashed line), and $h = 0$ (dotted line). $s = 0.001$, $u = 10^{-6}$, and $\lambda = 2$.

Equilibrium frequencies are then obtained by solving $\Delta x = 0$ (see APPENDIX A). The case for which W is the deleterious allele is straightforward because both gBGC and selection favor the S allele. The deleterious allele is maintained at a lower frequency than under mutation/selection equilibrium (see also BENGTSOON 1990):

$$x_{\text{eq}}^W \approx \frac{\lambda u}{(h + F - hF)s + b(1 - F)}. \quad (3)$$

The first case (with S being the deleterious allele), the one underlying the Achilles' heel hypothesis, is much less trivial. It leads to three selection regimes (Figures 1 and 2). If gBGC is weak relative to selection ($b \ll s$), the equilibrium frequency is

$$x_{\text{eq}}^S \approx \frac{u}{(h + F - hF)s - b(1 - F)}. \quad (4a)$$

The deleterious S allele is maintained at low frequency, but slightly higher than at mutation/selection equilibrium. On the contrary, if gBGC is strong ($b \gg s$), the deleterious allele is close to fixation:

$$x_{\text{eq}}^S \approx 1 - \frac{\lambda u}{(h - hF - 1)s + b(1 - F)}. \quad (4b)$$

Finally, if gBGC and selection are of similar intensity, the two forces can maintain the deleterious allele at intermediate frequency:

$$x_{\text{eq}}^S \approx \frac{b(1 - F) - (h + F - hF)s}{(1 - F)(1 - 2h)s}. \quad (4c)$$

The condition that ensures the existence of this last equilibrium point $0 \leq x_{\text{eq}}^S \leq 1$ is $(h + F - hF)s/(1 - F) \leq b \leq (1 - h + hF)s/(1 - F)$, and analysis of the derivative $(\partial \Delta x / \partial x)|_{x=x_{\text{eq}}^S}$ shows that this equilibrium is stable between the two limits (Figure 2). In panmixia, this case is fully equivalent to overdominance. Below a given selfing threshold, it is well known that the equilibrium is

FIGURE 2.—The three selection regimes as a function of b and s for different combinations of dominance levels (a, c, and e, $h = 0.1$; b, d, and f, $h = 0.3$) and mating systems (a and b, $F = 0$; c and d, $F = 0.5$; e and f, $F = 0.95$). 1: gBGC overwhelms selection and the GC deleterious allele goes to fixation. 2: gBGC and selection are of similar intensities and an overdominance-like equilibrium is reached. 3: Selection overwhelms gBGC and the deleterious allele becomes extinct.

not stable if overdominance is asymmetrical (KIMURA and OHTA 1971). However, in the present case, the equilibrium is always stable because equivalent selection parameters depend on F and the range of the overdominant-like domain narrows as inbreeding increases (Figure 2). In the special case of fully recessive alleles in panmixia ($F = 0$),

$$x_{eq}^S \approx \frac{b + \sqrt{b^2 + 4u(b+s)}}{2(b+s)}, \quad (4d)$$

which vanishes to the well-known equilibrium $x_{eq}^S = \sqrt{u/s}$ without gBGC (HALDANE 1937). As long as gBGC is small ($b < 0.1$), Equation 4d also holds for lethal alleles and gives results very close to those of BENGTSOON (1990); also A. POPA, S. GLEMIN, E. POPA, D. MOUCHIROUD, and C. GAUTIER, unpublished results.

Dynamics in finite populations and time properties:

In finite populations, the first two moments of the stationary probability distribution of allele frequency are required to compute the mean load and mean inbreeding depression (e.g., GLÉMIN *et al.* 2003). The work of NAGYLAKI (1983a) can be extended to in-

corporate the effect of dominance and inbreeding using the classical diffusion theory and the reparameterization given by (2). The WRIGHT (1937) stationary distribution is given by

$$\Phi(x) = Kx^{4(\alpha N/(1+F))u-1}(1-x)^{4(\alpha N/(1+F))\lambda u-1} \times e^{-2(\alpha N/(1+F))x(-b(1-F)+s(h+F-hF)+sx(1-F)(1-2h))} \quad (5a)$$

if S is the deleterious allele and

$$\Phi(x) = Kx^{4(\alpha N/(1+F))\lambda u-1}(1-x)^{4(\alpha N/(1+F))u-1} \times e^{-2(\alpha N/(1+F))x(b(1-F)+s(h+F-hF)+sx(1-F)(1-2h))} \quad (5b)$$

if W is the deleterious allele, where K is an integration constant, ensuring that $\int_0^1 \Phi(x)dx = 1$ and $N_e = \alpha(N/(1+F))$, where N is the effective population size of the equivalent population without inbreeding (POLLAK 1987; NORDBOG 1997), and α ($\alpha = 1$ in panmixia; $0 < \alpha \leq 1$ under inbreeding) is a coefficient introduced to summarize the additional reduction in effective population size due to hitchhiking and bottleneck effects associated with inbreeding (GLÉMIN 2007).

There is no general analytical solution for K . However, for the special case of genic selection in panmixia ($F = 0$ and $h = \frac{1}{2}$), the moments of Φ are

$$E_{\Phi}[x^i] = \frac{\Gamma(i + 4Nu)_1 F_1(i + 4Nu; i + 4Nu(1 + \lambda); -2N(s - 2b))}{\Gamma(4Nu)_1 F_1(4Nu; 4Nu(1 + \lambda); -2N(s - 2b))} \quad (6a)$$

if S is the deleterious allele and

$$E_{\Phi}[x^i] = \frac{\Gamma(i + 4N\lambda u)_1 F_1(i + 4N\lambda u; i + 4Nu(1 + \lambda); -2N(s + 2b))}{\Gamma(4N\lambda u)_1 F_1(4N\lambda u; 4Nu(1 + \lambda); -2N(s + 2b))}, \quad (6b)$$

if W is the deleterious allele, where Γ is the gamma function and ${}_1F_1$ is the regularized confluent hypergeometric function (ABRAMOWITZ and STEGUN 1970). Analytical expression cannot be given in the general case, and numerical integrations are used (see APPENDIX B).

The probability of fixation of a new mutation is also useful (see APPENDIX A). General analytical expressions can be obtained but they are formidable and not given here. However, the probability of fixation of a new mutation under genic selection in panmixia simply reduces to

$$P_{\text{fix}} = \frac{1 - e^{s-b}}{1 - e^{-2N(s-b)}} \approx \frac{s - b}{1 - e^{-2N(s-b)}} \quad (7a)$$

if S is the deleterious allele and

$$P_{\text{fix}} = \frac{1 - e^{s+b}}{1 - e^{-2N(s+b)}} \approx \frac{s + b}{1 - e^{-2N(s+b)}} \quad (7b)$$

if W is the deleterious allele. Approximations hold for $|s - b| \ll 1$ and $|s + b| \ll 1$, respectively (see also NAGYLAKI 1983a).

The last quantities of interest are sojourn and fixation times. Time properties are needed to evaluate whether,

on average, the allele life span is much shorter than the recombination hotspot life span. This is implicitly assumed by modeling a constant gBGC intensity. In the panmictic codominant case, the absorption time has been already derived by NAGYLAKI (1983a). Conversion only modifies the selection intensity (genic selection), and S deleterious alleles quickly lead to either extinction or fixation. The mean absorption time of a new mutant is always shorter than that under pure genetic drift (see Equations 28–33 in NAGYLAKI 1983a). On the contrary, as the deterministic analysis showed, gBGC can maintain recessive deleterious alleles at intermediate frequencies for a long time by mimicking overdominance. Conventional diffusion results can also be applied (*e.g.*, EWENS 2004) to compute the mean sojourn and fixation times of a new mutation (see APPENDIX A). In the general case, these expressions have no analytical expression and numerical integrations are performed (see APPENDIX B). However, under the “overdominance-like” regime, deleterious alleles rarely reach fixation. It is thus more interesting to compute the mean time until the mutant allele reaches the deterministic equilibrium frequency (x_{eq}) for the first time. I used the expression of KIMURA and OHTA (1973) that gives the mean time to reach a given frequency for the first time, starting from frequency $p_0 < x_{\text{eq}}$. For fully recessive mutations ($h = 0$) in panmixia, the equilibrium frequency is well approximated by $x_{\text{eq}} = b/(b + s)$ (Equation 4d with $s > b \gg u$). Once again, no analytical expression is available and numerical integrations are performed. Finally, under the overdominance-like regime, the allele can persist for a long time if it is not initially lost (*e.g.*, TAKAHATA 1990). To evaluate this conditional persistence, I also computed the mean sojourn time of an allele starting

TABLE 3
Time properties under the “overdominant-like” regime: effect of s and b ($N = 10,000$, $h = 0$)

	x_{eq}	Mean sojourn time	Fixation time	Mean time to reach x_{eq}	Mean sojourn time from x_{eq}
$s = 0.0005$					
$b = 0$	—	53	26,846	—	—
$b = 0.0001$	0.2	67	40,864	4,383	$1 \times E + 08$
$b = 0.0002$	0.4	105	73,314	9,795	$7 \times E + 08$
$b = 0.0003$	0.6	95	73,314	15,219	$1 \times E + 08$
$b = 0.0004$	0.8	72	40,864	19,003	390,270
$s = 0.001$					
$b = 0$	—	50	18,200	—	—
$b = 0.0002$	0.2	74	34,008	4,447	$1 \times E + 08$
$b = 0.0004$	0.4	313	145,154	9,322	$4 \times E + 09$
$b = 0.0006$	0.6	208	145,154	11,355	$1 \times E + 08$
$b = 0.0008$	0.8	88	34,008	12,286	34,008
$s = 0.002$					
$b = 0$	—	47	12,205	—	—
$b = 0.0004$	0.2	101	33,633	4,494	$2 \times E + 08$
$b = 0.0008$	0.4	4,593	$1 \times E + 06$	7,220	$3 \times E + 11$
$b = 0.0012$	0.6	2,472	$1 \times E + 06$	7,069	$1 \times E + 08$
$b = 0.0016$	0.8	137	33,633	7,451	26,182

TABLE 4

Time properties under the “overdominant-like” regime: effect of N ($h = 0$, $s = 0.001$, $b = 0.0002$, $x_{eq} = 0.2$)

	Mean sojourn time without gBGC	Mean sojourn time	Fixation time	Mean time to reach x_{eq}	Mean sojourn time from x_{eq}
$N = 1,000$	43	44	4,385	431	706,085
$N = 5,000$	49	63	20,431	2,191	$3 \times E + 07$
$N = 10,000$	50	74	34,007	4,447	$1 \times E + 08$
$N = 50,000$	54	466	419,015	20,230	$4 \times E + 10$
$N = 100,000$	56	14,977	$1 \times E + 07$	28,752	$4 \times E + 13$
$N = 500,000$	61	$5 \times E + 17$	$4 \times E + 20$	42,339	$4 \times E + 40$
$N = 1,000,000$	63	$8 \times E + 34$	$7 \times E + 37$	47,925	$1 \times E + 75$

from frequency x_{eq} . The numerical results are presented in Tables 3 and 4].

On average, gBGC makes S deleterious alleles persist longer in populations. However, the time to reach the deterministic equilibrium frequency for the first time is still short. Therefore, computing the equilibrium frequency on the basis of constant b is reliable despite the short life span of recombination hotspots. On the contrary, once the allele reaches the equilibrium frequency it will persist for a very long time. If it is not initially lost, a deleterious S allele will thus persist in the population over the hotspot life span.

Impacts of gBGC on the mutation load: I now consider how gBGC affects the mutation load. The load is defined as

$$L \equiv 1 - \bar{W} = (F + 2h - 2hF)sx + (1 - F)(1 - 2h)sx^2. \quad (8)$$

Here we need to consider both sites for which either S or W alleles are deleterious, in proportions q and $1 - q$, respectively, so that

$$L = qL_S + (1 - q)L_W. \quad (9)$$

For nonsynonymous positions and noncoding functional regions, it seems reasonable to assume that $q = \frac{1}{2}$. However, for synonymous positions under selection for codon usage, q will depend on the distribution of preferred codons. As GC ending codons are often more preferred than AT ending ones (DURET and MOUCHIROUD 1999; WANG and ROOSSINCK 2006), q can be close to 0.

gBGC and the load structure in panmictic populations: In infinite panmictic populations, without gBGC, the mutation load is equal to twice the mutation rate (HALDANE 1937). In finite populations, deterministic results hold for strongly deleterious mutations ($s \gg 1/N_e$). Slightly deleterious alleles ($\sim s \leq 1/N_e$) can drift to fixation, causing a much higher load (“fixation load”) than that caused by strongly deleterious mutations segregating at low frequency (“segregating load”), *i.e.*, $\sim L \approx s$ (Figure 3a and see KIMURA *et al.* 1963). Conse-

quently, the relative part of the fixation load (or “drift load,” see the distinction below) and the segregating load depends on both the effective population size and the distribution of selection coefficients.

gBGC affects both the magnitude and the structure of the mutation load. In infinite populations, and more generally for strongly deleterious alleles ($N_e s \gg 1$), the deterministic load is obtained by replacing x by x_{eq} , given by Equations 4 in Equations 8 and 9:

$$L \approx 2uhs \left(\frac{q}{hs - b} + \frac{\lambda(1 - q)}{hs + b} \right) \quad \text{if } b \ll hs \quad (10a)$$

$$L \approx \frac{q(b^2 - h^2 s^2)}{s(1 - 2h)} \quad \text{if } hs < b < (1 - h)s \quad (10b)$$

$$L \approx qs \quad \text{if } b \gg s. \quad (10c)$$

Under the first condition ($b \ll hs$), gBGC increases the load if $b > hs(1 - 2q/(q + \lambda - q\lambda))$. Under the other conditions, the load due to W deleterious mutations is negligible and gBGC always increases the load. Using $\lambda = 2$ and $q = \frac{1}{2}$, gBGC increases the load due to mutations for which $hs < 3b$. Given that b is quite low (SPENCER *et al.* 2006), gBGC increases the load only due to rather weakly deleterious mutations.

In finite populations, the load must be integrated over the Φ distribution, which leads to

$$L = (F + 2h - 2hF)sE_\Phi[x] + (1 - F)(1 - 2h)sE_\Phi[x^2]. \quad (11)$$

$E_\Phi[x]$ and $E_\Phi[x^2]$ can be computed numerically (APPENDIX B) and introduced in (11).

For genic selection ($h = \frac{1}{2}$) in panmixia ($F = 0$), the approximations

$$L_W \approx s \frac{1}{1 + e^{2N_e(s+2b)}/\lambda} + \frac{2s\lambda u}{s + 2b} \left(1 - \frac{1}{1 + e^{2N_e(s+2b)}/\lambda} \right) \quad (12a)$$

and

FIGURE 3.—The load ($\times 10^6$) as a function of s for different effective population sizes (thick lines, $N = 5000$; thin lines, $N = 10,000$; dashed lines, $N = 50,000$; dotted lines, $N = 100,000$), without (a) or with (b–d) gBGC ($b = 0.0002$). The straight lines correspond to equilibrium expectations. (b) Load due to GC deleterious alleles. (c) Load averaged over half GC and half AT deleterious alleles. (d) Load averaged over 10% of GC deleterious alleles and 90% of AT deleterious alleles with a bias in favor of AT alleles. This last case would correspond to the load caused by third codon position mutations under codon usage selection, mainly in favor of GC alleles. $h = 0.5$, $u = 10^{-6}$, and $\lambda = 2$.

$$L_S \approx s \frac{1}{1 + \lambda e^{2N_e(s-2b)}} + L_{\text{det}} \left(1 - \frac{1}{1 + \lambda e^{2N_e(s-2b)}} \right) \quad (12b)$$

can be obtained (see APPENDIX C), where L_{det} is the deterministic load given by Equations 10. Interestingly, these approximations are very robust and hold for a wide range of parameters (see APPENDIX D for numerical validation). The analysis of Equations 12 and numerical explorations show that (i) if $N_e b < 1$, the effect of gBGC is negligible (but see below) and the mutations maximizing the load (either W or S) are of the order of $s \approx 1/N_e$, and (ii) if $N_e b > 1$, the mutations maximizing the load are S mutations with $s \approx 2b$ (see Figure 3 and APPENDIX C). gBGC increases the fraction of the fixation load beyond the drift load *sensu stricto*, *i.e.*, the load due to deleterious mutations fixed by drift only.

Minimum load: As explained above, if mutations are biased toward W alleles, under certain circumstances, gBGC can decrease the load (see APPENDIX C). A minimum is thus reached for low gBGC levels. In

panmictic populations, the gBGC level that minimizes the segregating load is

$$b_{\min}^{\text{seg}} \approx hs \left(1 - \frac{2}{1 + \sqrt{\lambda}} \right). \quad (13)$$

While gBGC can greatly increase the fixation load, a very low gBGC level can also reduce it, and the gBGC level that minimizes the fixation load is

$$b_{\min}^{\text{fix}} \approx \frac{\ln(\lambda)}{4N_e}. \quad (14)$$

Both expressions clearly show that gBGC can minimize the total load only if $\lambda > 1$; *i.e.*, the mutation is biased toward W alleles. Interestingly, the gBGC level that minimizes the fixation load does not depend on s but instead only on N_e and λ . Since (13) depends on h and s , it is hard to predict which gBGC level will minimize the load for a given distribution of h and s . However, gBGC will greatly increase the fixation load as far as $b > \ln(\lambda)/2N_e$ (see APPENDIX C). Since $b_{\min}^{\text{fix}} < b_{\min}^{\text{seg}}$ in

most cases, $b_{\min}^{\text{fix}}$ is thus a very good approximation of the gBGC level that minimizes the total load, whatever the distribution of mutational effects is. This rationale was confirmed by numerical analyses (APPENDIX D).

Quantitative effects of gBGC on the load: To quantitatively assess the theoretical effects predicted above, I computed the total load in a genome experiencing a 3% gBGC hotspot with $b = 0.0002$ on average [according to estimates in humans (SPENCER *et al.* 2006), and see also GALTIER *et al.* 2009] and exposed to a flow of deleterious mutations with a gamma-shaped distribution of selection coefficients with mean $s_{\text{mean}} = 0.0325$ and shape parameter $= 0.23$ [according to estimates in humans (EYRE-WALKER *et al.* 2006), and see also GALTIER *et al.* 2009]. For simplicity, fitness is assumed to be multiplicative across loci. However, recombination and conversion are not distributed similarly in all organisms. For instance, recombination is probably not localized in hotspots in *Caenorhabditis elegans* (ROCKMAN and KRUGLYAK 2009). I thus also tested more homogenous and more heterogeneous gBGC distributions. To make the different cases comparable, I based my computations on a constant total amount of gBGC ($b_{\text{total}} = 0.0002 \times 0.03 = 6 \times 10^{-6}$). In addition to the first case (3% hotspots with $b = 0.0002$), I assumed that gBGC was homogeneously distributed over the whole genome, distributed over half of the genome, and in 0.3% hotspots with $b = 0.002$. For simplicity, all mutations are assumed to be codominant ($h = \frac{1}{2}$), and the genomic mutation rate is set at $U = 0.01$ (see APPENDIX B). The results are presented in Table 5. They show that in rather large populations, gBGC hotspots as low as 3% can increase the load two- or threefold. Even in small populations ($N = 10,000$, similar to effective size in humans), rare gBGC hotspots can increase the load by almost 10%. However, this analysis also shows that weak gBGC throughout the genome has a weak effect or can even reduce the load by compensating for the mutational bias. This occurs when $b < \ln(\lambda)/2N_e$, as predicted above.

Impacts of gBGC on inbreeding depression: Inbreeding depression is defined as the reduction in fitness of selfed (and more generally inbred) individuals compared to outcrossed individuals,

$$\delta \equiv 1 - \frac{\bar{W}_{\text{self}}}{\bar{W}_{\text{out}}} = \frac{s(1-F)(1-2h)x(1-x)}{1-2x(1-x)(1-F)hs - (x^2 + Fx(1-x))s} \approx s(1-F)(1-2h)x(1-x), \quad (15)$$

where $\bar{W}_{\text{self}}$ and $\bar{W}_{\text{out}}$ are the mean fitness of outcrosses and selfcrosses, respectively (CHARLESWORTH and CHARLESWORTH 1987; CHARLESWORTH and WILLIS 2009). The approximation is very good in most conditions, because $\bar{W}_{\text{out}} \approx 1$ under weak ($s \ll 1$) and strong selection ($x \ll 1$) (see GLÉMIN *et al.* 2003). Similar to the load, considering both sites for which either S or W alleles are deleterious, in proportion q and $1 - q$, respectively, we get

TABLE 5
Genomic load at mutation–selection–drift equilibrium contributed by 10^6 loci with $u = 10^{-8}$ (genomic mutation rate $U = 0.01$) and $\lambda = 2$, and s follows a gamma distribution with mean 0.0325 and shape 0.23

	Load			
	Without gBGC	Homogenous gBGC	Hotspots 1	Hotspots 2
$N = 10,000$				
$F = 0$	0.634	0.615	0.689	0.742
$F = 0.2$	0.637	0.624	0.666	0.706
$F = 0.5$	0.649	0.642	0.655	0.678
$F = 0.7$	0.670	0.667	0.670	0.682
$F = 0.95$	0.854	0.854	0.854	0.854
$F = 0.99$	0.993	0.993	0.993	0.993
$N = 50,000$				
$F = 0$	0.143	0.138	0.301	0.399
$F = 0.2$	0.142	0.131	0.240	0.309
$F = 0.5$	0.144	0.135	0.183	0.219
$F = 0.7$	0.150	0.144	0.165	0.185
$F = 0.95$	0.240	0.239	0.240	0.242
$F = 0.99$	0.496	0.496	0.496	0.496
$N = 100,000$				
$F = 0$	0.074	0.096	0.248	0.351
$F = 0.2$	0.071	0.074	0.181	0.253
$F = 0.5$	0.071	0.065	0.117	0.153
$F = 0.7$	0.073	0.068	0.093	0.111
$F = 0.95$	0.115	0.114	0.115	0.118
$F = 0.99$	0.257	0.257	0.257	0.257

Homogenous gBGC, the entire genome is affected by gBGC with $b = 6 \times 10^{-6}$; with hotspots 1, 3% of the genome belongs to conversion hotspots with $b = 2 \times 10^{-4}$ (similar to human data); with hotspots 2, 0.3% of the genome belongs to conversion hotspots with $b = 2 \times 10^{-3}$.

$$\delta = q\delta_s + (1 - q)\delta_w. \quad (16)$$

gBGC and the genetic basis of inbreeding depression in panmictic populations: In infinite panmictic populations without gBGC, inbreeding depression depends only on mutation rates and dominance levels. Partially recessive mutations ($h < \frac{1}{2}$) contribute only to inbreeding depression, and the more recessive they are, the higher the inbreeding depression (CHARLESWORTH and CHARLESWORTH 1987). In finite populations, deterministic results hold for strongly deleterious mutations ($s \gg 1/N_e$), which contribute mostly to inbreeding depression. Contrary to the load, weakly deleterious mutations ($\sim s \leq 1/N_e$) contribute little to inbreeding depression (Figure 4, a and c, and see BATAILLON and KIRKPATRICK 2000).

Like the load, gBGC affects both the magnitude and the structure of inbreeding depression. In infinite populations, and more generally for strongly deleterious alleles ($N_e s \gg 1$), replacing x by x_{eq} given by Equations 4 in Equations 15 and 16 leads to

FIGURE 4.—Inbreeding depression ($\times 10^6$) as a function of s without (a and c) or with (b and d) gBGC ($b = 0.0002$). (a and b) $h = 0.2$: thick lines, $N = 5000$; thin lines, $N = 10,000$; dashed lines, $N = 50,000$; dotted lines, $N = 100,000$. (c and d) $N = 10,000$: thick lines, $h = 0.4$; thin lines, $h = 0.2$; dashed lines, $h = 0.1$; dotted lines, $h = 0.05$. $u = 10^{-6}$, $\lambda = 2$.

$$\delta \approx \frac{u(1-2h)s}{2} \left(\frac{q}{hs-b} + \frac{\lambda(1-q)}{hs+b} \right) \quad \text{if } b \ll hs \quad (17a)$$

$$\delta \approx q \frac{(b-hs)((1-h)s-b)}{2s(1-2h)} \quad \text{if } hs < b < (1-h)s \quad (17b)$$

$$\delta \approx \frac{\lambda u(1-2h)s}{2} \left(\frac{q}{b-(1-h)s} + \frac{1-q}{b+hs} \right) \quad \text{if } b \gg s. \quad (17c)$$

The effect of gBGC on inbreeding depression is not monotonic. Like the load, gBGC increases inbreeding depression if $b > hs(1 - 2q/(q + \lambda - q\lambda))$. However, contrary to the load, a strong gBGC decreases inbreeding depression, which tends to 0 as b increases, while the load tends to qs (Equation 10c). An analysis of Equation 17b shows that mutations that maximize inbreeding depression are those that also maximize the load, *i.e.*, S deleterious mutations with $s \approx 2b$.

In finite populations, inbreeding depression must be integrated over the Φ distribution, which leads to

$$\delta \approx s(1-F)(1-2h)(E_\Phi[x] - E_\Phi[x^2]) \quad (18)$$

(see also GLÉMIN *et al.* 2003). While it is not possible to get an analytical expression of (18), numerical computations (see APPENDIX B) show that S deleterious

mutations with $s \approx 2b$ also maximize inbreeding depression in finite populations (Figure 4). More broadly, inbreeding depression is maximal under the overdominant-like selection regime (gray area in Figure 2). Once again, even low to moderate gBGC markedly affects the genetic structure of inbreeding depression. First, mutations of intermediate effects contribute the most to inbreeding depression, *i.e.*, up to one order of magnitude higher than strongly deleterious mutations (compare Figure 4a with 4b). Second, even nearly additive mutations can have a substantial effect (compare Figure 4c with 4d).

Since little is known about the distribution of dominance coefficients, especially the dominance of mildly deleterious mutations (of the order of b), it is difficult to quantitatively predict the full impact of gBGC on inbreeding depression. We can conclude that, on average, gBGC should increase inbreeding depression. However, further insight into mutational parameters is crucial to assess the quantitative impact of gBGC.

Joint effect of gBGC and mating system on the load and inbreeding depression: Selfing, or more generally inbreeding, slightly reduces the segregating load through the purging of recessive mutations (OHTA and COCKERHAM 1974), but can substantially increase the fixation load because of the effective population size reduction under inbreeding: $N_e = \alpha(N/(1+F))$ (see above and POLLAK 1987; NORDBOG 1997; GLÉMIN

FIGURE 5.—Effective population size (a and b) and the load ($\times 10^6$) (c–f) as a function of F for different gBGC intensities (thick lines, $b = 0$; thin lines, $b = 0.0001$; dashed lines, $b = 0.0002$; dotted lines, $b = 0.0005$). The effective population size depends on F under the background selection (BS) model (CHARLESWORTH *et al.* 1993), using Equations 16 and 17 in GLÉMIN (2007): $N_e(F) = N \exp[U(1/(h_d s_d + R) - 1/((h_d + F - h_d F)s_d + R(1 - F)))]$, where U is the genomic deleterious mutation rate, R is the genomic recombination rate, s_d is the mean selection coefficient against strongly deleterious mutations, and h_d is their dominance coefficient. $N = 10,000$, $U = 0.2$, $h_d = 0.1$, and $s_d = 0.05$. (a, c, and e) $R = 5$, “weak” BS; (b, d, and f) $R = 0.5$, “strong” BS. (c and d) Load averaged over half GC and half AT deleterious alleles, with a bias in favor of AT alleles. (e and f) Load averaged over 10% of GC deleterious alleles and 90% of AT deleterious alleles with a bias in favor of AT alleles; see Figure 3. $h = 0.5$, $u = 10^{-6}$, and $\lambda = 2$.

2007). In numerical examples, I assumed that α decreases with F according to the background selection model (CHARLESWORTH *et al.* 1993; NORDBERG *et al.* 1996), as in GLÉMIN (2007). With gBGC, selfing thus has two opposite effects on the fixation load. Selfing increases the drift load *sensu stricto* but decreases the fixation load due to gBGC. A surprising consequence is that the load can be higher in outcrossing than in selfing populations (Figure 5). Quantitatively this is also expected, even with a gBGC hotspot affecting just 3% of the genome (Table 5). Another consequence is that a minimum load can be reached for intermediate selfing rates (Figure 5 and Table 5).

Generally, the effect of selfing is simpler for inbreeding depression. Purging, N_e reduction, and suppression of gBGC contribute to decreasing inbreeding depression in selfing populations (Figure 6a). However, there are special cases in which maximum inbreeding depression is reached for intermediate selfing rates (Figure 6b). In such cases, in outcrossing populations, gBGC is strong enough to sweep polymorphism out and reduce inbreeding depression ($b > s$, regime 1 in Figure 2). As the selfing rate increases, gBGC declines, and the selection dynamics become overdominant-like (regime 2,

Figure 2), thus maximizing inbreeding depression. For high selfing rates, gBGC vanishes (regime 3 in Figure 2) and deleterious alleles are either purged or fixed if there is substantial drift. This is similar to the effect of selfing on inbreeding depression caused by asymmetrical overdominance, where inbreeding depression also peaks for intermediate selfing rates (ZIEHE and ROBERTS 1989; CHARLESWORTH and CHARLESWORTH 1990). In the present case, the range of parameters leading to this peculiar behavior is narrow because the overdominant-like region depends on the selfing rates and can vanish either for low or for high selfing rates (Figure 2).

DISCUSSION

I analyzed the effect of gBGC on the fate and consequences of deleterious mutations by extending previous theoretical studies to also include population processes, such as drift and inbreeding, and the whole range of mutations, including direction ($W \rightarrow S$ and $S \rightarrow W$), strength, and dominance. gBGC affects S and W deleterious alleles in opposite ways. However, except in special cases, these opposite effects do not cancel out

FIGURE 6.—Inbreeding depression ($\times 10^6$) as a function of F for different gBGC intensities (thick lines, $b = 0$; thin lines, $b = 0.0001$; dashed lines, $b = 0.0002$; dotted lines, $b = 0.0005$). Inbreeding depression is averaged over half GC and half AT deleterious alleles. The effective population size depends on F as in Figure 5 (same parameters). (a) $s = 0.002$; (b) $s = 0.0005$; (c) $s = 0.0002$. $h = 0.2$, $u = 10^{-6}$, and $\lambda = 2$.

because S alleles are much more strongly affected than W alleles. Moreover, while significant gBGC levels are likely limited to recombination hotspots [only 3% of the genome in humans (MYERS *et al.* 2005)], I showed that both the quantitative and the qualitative effects of gBGC on the load and inbreeding depression are far from being negligible.

Deleterious impacts of gBGC: Segregation of deleterious alleles and genetic diseases: Depending on selection and dominance levels, gBGC can slightly increase the frequency of deleterious S alleles before their elimination, drive them to fixation, or maintain them at intermediate frequencies for long periods of time in an

overdominant-like way (Figure 2). In particular, the frequency of recessive alleles, even strongly deleterious ones, can be markedly increased by gBGC. Even moderate gBGC intensities can thus significantly affect lethal alleles. For instance, taking $u = 10^{-8}$ and $b = 0.0002$ as estimated in human recombination hotspots (MYERS *et al.* 2005), the frequency of a lethal allele can be more than doubled (see Equation 4d). On the contrary, gBGC helps in purging W deleterious alleles. We thus expect that (i) genes near recombination hotspots are more likely involved in genetic diseases and (ii) S alleles contribute disproportionately to genetic diseases, especially those found in appreciable frequencies. These predictions were recently confirmed through a genome-wide survey of disease-related mutations (A. NECSULEA, A. POPA, D. N. COOPER, P. D. STENSON, D. MOUCHIROUD, C. GAUTIER, and L. DURET, unpublished results).

However, this kind of disease allele will be probably hard to identify individually through population genomic approaches. While formally equivalent in a one-locus model, recessive deleterious alleles balanced by gBGC and “true” balancing selection should leave different signatures at the molecular level. Because conversion tracts are likely only a few hundred base pairs long and mainly occur in recombination hotspots, only very short haplotypes are expected to be dragged up at intermediate frequency. It thus should be difficult to detect such alleles maintained in the population for a long time, contrary to other balancing selection mechanisms (CHARLESWORTH 2006a).

Fraction of the load due to gBGC: Through its effect on the fate of deleterious alleles, gBGC affects both the magnitude and the architecture of the genetic load (Figure 7). Under realistic conditions of mutation bias toward AT ($\lambda > 1$), gBGC slightly decreases the segregating load due to strongly deleterious mutations, but markedly increases the fixation load due to weakly deleterious mutations. For S mutations, the limit between the two mutation categories is mainly independent of the effective population size but depends on the intensity of gBGC. On the contrary, W mutations behave mainly as without gBGC and the limit between the two mutation categories is still $\sim s \approx 1/2N_e$. Quantitatively, the effect of gBGC on the segregating load is rather weak. The main effect of gBGC is thus to increase the fixation load by the proportion of mutations spanning $\sim 1/2N_e < s < 2b$. The paradoxical consequences are that (i) the fraction of the fixation load induced by gBGC becomes roughly independent of population size and (ii) the proportion of the load due to gBGC increases with effective population size.

Quantitatively, gBGC can have a significant effect. Assuming only 3% the genome experiencing gBGC hotspots of average intensity $b = 0.0002$, gBGC can increase the load by two- to threefold in large populations (Table 5). In humans (using $N = 10,000$), gBGC

FIGURE 7.—Effect of gBGC on the load structure. The gamma-shaped curve represents the distribution of selection coefficients, s . Deleterious mutations can either segregate in low frequency (open part), go to fixation because of gBGC (shaded part), or go to fixation because of drift (solid part). Compared to classical predictions without gBGC, gBGC adds the shaded part. Drift load *sensu stricto* and segregating load are also affected by gBGC but very weakly. In small populations (a), the drift load is higher and the gBGC fixation load lower than in large populations (b).

could contribute to 10% of the load. However, this is only a rough estimate based on selection parameters estimated from coding regions. On one hand, recombination seems to be depressed within genes (MYERS *et al.* 2005; COOP *et al.* 2008), which should reduce the load due to gBGC. For instance, if gBGC is twofold lower within genes than around them ($b = 0.0001$), only 5% of the load would be due to gBGC (data not shown). On the other hand, noncoding functional regions can also contribute to the load, especially in gene control regions (KEIGHTLEY *et al.* 2005; EORY *et al.* 2010), where recombination can be high (COOP *et al.* 2008). Despite these uncertainties, gBGC may have fundamental fitness implications, not merely practical consequences for the detection of selection at the molecular level (BERGLUND *et al.* 2009; GALTIER *et al.* 2009). These results confirm the pertinence of the “genomic Achilles’ heel” metaphor proposed by GALTIER and DURET (2007).

Fixation load and heterosis: The load structure is also crucial for predicting heterosis, the increase in fitness of crosses between populations or lines compared to parents. Heterosis is created and proportional to among-population variance in deleterious allele frequencies (WHITLOCK *et al.* 2000; GLÉMIN *et al.* 2003). Without gBGC, mutations contributing to heterosis also contribute to the drift load, and heterosis is mainly

expected in small and isolated populations (WHITLOCK *et al.* 2000; GLÉMIN *et al.* 2003). As discussed above, gBGC generates the fixation load in large populations, especially for $s \sim 2b$ (see above). In addition, under these conditions, the variance in allele frequencies is also high. gBGC compensates for selection and deleterious alleles behave almost neutrally. The variance in allele frequency is thus roughly that expected under mutation–drift equilibrium, $V_{\Phi}[x] = \lambda / (1 + \lambda)^2 (1 + 4Nu(1 + \lambda))$ (WRIGHT 1937), which can be high, even in large populations. This is in sharp contrast with the low variance expected without gBGC, $V_{\Phi}[x] \approx u / hs (1 + 4Nhs)$ (BATAILLON and KIRKPATRICK 2000). While this rationale should be confirmed by detailed analyses of the interaction of selection and gBGC in subdivided populations, it leads to the surprising conclusion that gBGC could create heterosis in large populations.

Minimum load and the evolution of gBGC and recombination landscapes: Although gBGC may have deleterious fitness consequences, it is surprising that it evolved in many taxa (Table 1 and reviewed in DURET and GALTIER 2009a). BIRDELL (2002) initially suggested that gBGC may have evolved as a response to mutational bias toward AT ($\lambda > 1$, here). Indeed, I show that a minimum load is reached for weak gBGC ($b \approx \ln(\lambda) / 4N$, Equation 14). This result is very general whatever the distribution of fitness effects of mutations (APPENDIX D). However, the range of optimal gBGC is narrow, and gBGC increases the load as far as $b > \ln(\lambda) / 2N$ (APPENDIX C). In humans, using $N = 10,000$ and $\lambda = 2$, gBGC levels that minimize the load are $\sim 1.17 \times 10^{-5}$, *i.e.*, one order of magnitude lower than the average bias observed in recombination hotspots (MYERS *et al.* 2005). However, selection on conversion modifiers will not necessarily minimize the load because of gametic disequilibrium generated between modifiers and fitness loci (BENGTTSSON and UYENOYAMA 1990). Selection for limitation of somatic AT-biased mutations could also have selected for GC-biased mismatch repair machinery (BROWN and JIRICNY 1987). If the bias level that would be selected for somatic reasons is $> \ln(\lambda) / 2N$, a side effect would be the generation of a substantial load at the population level. Finally, it is interesting to note that when synonymous codon positions are under selection for translation accuracy, optimal gBGC levels can be higher than gBGC levels that minimize the protein load, especially when most optimal codons end in G or C ($q > \frac{1}{2}$).

Conversely, gBGC could also affect the evolution of recombination landscapes, which could evolve to reduce the gBGC load. Surprisingly, for a given recombination/conversion level, the hotspot distribution does not appear to be optimal (Table 5). Homogenous but lower gBGC has little effect on the load and could even reduce it by offsetting the mutational bias, provided that $b < \ln(\lambda) / 2N$. However, in species where recombination mainly occurs at hotspots, such as in

humans, mice, and yeasts (NISHANT and RAO 2005), one can speculate that the hotspot localization outside genes could be a response to avoid the deleterious effects of gBGC.

Up to now, these verbal arguments have not been assessed theoretically (but see BENGTTSSON and UYENOYAMA 1990 for a different kind of conversion bias). Population genetics models are necessary to test these hypotheses concerning the evolution of gBGC and recombination landscapes and to pinpoint the key parameters that might govern their evolution.

gBGC and the evolution of mating systems: Deleterious mutations also play a crucial role in the evolution of mating systems. They are the main source of inbreeding depression, which balances the automatic advantage of selfing. The drift load is also thought to contribute to the extinction of selfing species. Since they are mainly homozygous, selfing species are mostly free from gBGC and its deleterious impacts. I discuss below how this might affect the evolution of mating systems.

Inbreeding depression and the shift in mating systems: Inbreeding depression plays a key role in the evolution of mating systems (CHARLESWORTH and CHARLESWORTH 1987; CHARLESWORTH 2006b). Since it balances the automatic advantage of selfing, high inbreeding depression favors outcrossing, while selfing can evolve when it is low. Moreover, selfing helps to purge strongly deleterious mutations, thus decreasing inbreeding depression. This positive feedback reinforces the disruptive selection on the selfing rate and prevents the transition from selfing to outcrossing (LANDE and SCHEMSKE 1985).

Theoretical results suggest that, in most conditions, gBGC would reinforce inbreeding depression in outcrossing populations (Figure 6), which would prevent the evolution of selfing. In reverse, if selfing is initially selected for, recurrent selfing would reduce the load through both purging and avoidance of gBGC. Under this scenario, gBGC would reinforce disruptive selection on mating systems. However, under some conditions (see Figure 6), inbreeding depression peaks at intermediate selfing rates, as observed for asymmetrical overdominance (ZIEHE and ROBERDS 1989; CHARLESWORTH and CHARLESWORTH 1990). In theory, this could prevent the shift toward complete selfing and maintain stable mixed mating systems (CHARLESWORTH and CHARLESWORTH 1990; UYENOYAMA and WALLER 1991). However, this pattern is observed under restrictive conditions and it is very unlikely on the whole-genome scale. Dominance patterns are crucial for predicting inbreeding depression, especially with gBGC. Contrary to the load, it is thus difficult to evaluate the quantitative impact of gBGC on inbreeding depression. However, increased inbreeding depression in outcrossing species subject to gBGC seems to be the most likely scenario.

gBGC and the long-term evolution of mating systems: In the long term, the gBGC-induced load also challenges the “dead-end hypothesis,” which posits that, because of the reduction of selection efficacy, self-fertilizing species would accumulate weakly deleterious mutations in the long term, eventually leading to extinction (TAKEBAYASHI and MORRELL 2001). Because of gBGC, not drift, outcrossing species could also accumulate a load of weakly deleterious mutations (Figure 7), and they could suffer from a higher load than highly self-fertilizing species (Table 4). In line with this hypothesis, HAUDRY *et al.* (2008) found that in two outcrossing grass species, but not in two self-fertilizing ones, the d_N/d_S ratio is significantly higher for genes exhibiting GC enrichment. They speculated that substitutions in these genes might contribute to increasing the load in these two outcrossing grass species. Such results are still very sparse. In plants, evidence of strong gBGC is mainly restricted to grasses (but see WRIGHT *et al.* 2007). It will be necessary to conduct more in-depth studies to assess the phylogenetic distribution of gBGC in plants and other hermaphrodite organisms and to further test the genomic Achilles’ heel hypothesis in relation to mating systems. While theoretically possible, the quantitative effect of gBGC on the evolution of mating systems remains a new, open, and challenging question.

Conclusion: I showed that the interaction between gBGC and selection might have surprising qualitative consequences on load and inbreeding depression patterns. Given the few quantitative data available on gBGC levels and selection intensities (mainly in humans), it turns out that even weak genome-wide gBGC can have significant fitness impacts. gBGC should be taken into account not only for sequence analyses (BERGLUND *et al.* 2009; GALTIER *et al.* 2009), but also for its potential fitness consequences, for instance concerning genetic diseases. Interferences between gBGC and selection also give rise to new questions on the evolution of mating systems. However, most of the challenging conclusions given here have yet to be quantitatively evaluated. Quantification of gBGC and its interaction with selection in various organisms will be crucial in the future.

I thank N. Galtier, L. Duret, A. Poppa, and C. Gautier for discussions and two anonymous reviewers for helpful insights and suggestions to improve the manuscript. I also thank L. Duret and A. Necseulea for sharing a manuscript before publication. This is manuscript Institut des Sciences de l’Evolution de Montpellier ISEM 2010-029. This work was supported by the French Centre National de la Recherche Scientifique and Agence Nationale de la Recherche (ANR-08-GENM-036-01).

LITERATURE CITED

- ABRAMOWITZ, M., and I. A. STEGUN, 1970 *Handbook of Mathematical Functions*. Dover, New York.
- BATAILLON, T., and M. KIRKPATRICK, 2000 Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genet. Res.* **75**: 75–81.
- BENGTTSSON, B. O., 1990 The effect of biased conversion on the mutation load. *Genet. Res.* **55**: 183–187.

- BENGTTSSON, B. O., and M. K. UYENOYAMA, 1990 Evolution of the segregation ratio: modification of gene conversion and meiotic drive. *Theor. Popul. Biol.* **38**: 192–218.
- BERGLUND, J., K. S. POLLARD and M. T. WEBSTER, 2009 Hotspots of biased nucleotide substitutions in human genes. *PLoS Biol.* **7**: e26.
- BIRDELL, J. A., 2002 Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol. Biol. Evol.* **19**: 1181–1197.
- BROWN, T. C., and J. JIRICNY, 1987 A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* **50**: 945–950.
- BULMER, M., 1991 The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**: 897–907.
- CABALLERO, A., and W. G. HILL, 1992 Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics* **131**: 493–507.
- CHARLESWORTH, B., 1992 Evolutionary rates in partially self-fertilizing species. *Am. Nat.* **140**: 126–148.
- CHARLESWORTH, B., and D. CHARLESWORTH, 1998 Some evolutionary consequences of deleterious mutations. *Genetica* **102/103**: 3–19.
- CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- CHARLESWORTH, D., 2006a Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* **2**: e64.
- CHARLESWORTH, D., 2006b Evolution of plant breeding systems. *Curr. Biol.* **16**: R726–R735.
- CHARLESWORTH, D., and B. CHARLESWORTH, 1987 Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* **18**: 237–268.
- CHARLESWORTH, D., and B. CHARLESWORTH, 1990 Inbreeding depression with heterozygote advantage and its effect on selection for modifiers changing the outcrossing rate. *Evolution* **44**: 1469–1489.
- CHARLESWORTH, D., and J. H. WILLIS, 2009 The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**: 783–796.
- COOP, G., X. WEN, C. OBER, J. K. PRITCHARD and M. PRZEWSKI, 2008 High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**: 1395–1398.
- CROW, J. F., 1993 Mutation, mean fitness, and genetic load, pp. 3–42 in *Oxford Surveys in Evolutionary Biology*, Vol. 9, edited by D. FUTUYMA and J. ANTONOVICS. Oxford University Press, Oxford.
- DURET, L., and P. F. ARNDT, 2008 The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* **4**: e1000071.
- DURET, L., and N. GALTIER, 2009a Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**: 285–311.
- DURET, L., and N. GALTIER, 2009b Comment on “Human-specific gain of function in a developmental enhancer”. *Science* **323**: 714 (author reply: 714).
- DURET, L., and D. MOUCHIROUD, 1999 Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **96**: 4482–4487.
- DURET, L., A. EYRE-WALKER and N. GALTIER, 2006 A new perspective on isochore evolution. *Gene* **385**: 71–74.
- DURET, L., J. COHEN, C. JUBIN, P. DESSEN, J. F. GOUT *et al.*, 2008 Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline. *Genome Res.* **18**: 585–596.
- EORY, L., D. L. HALLIGAN and P. D. KEIGHTLEY, 2010 Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. *Mol. Biol. Evol.* **27**: 177–192.
- EWENS, W. J., 2004 *Mathematical Population Genetics. I. Theoretical Introduction*. Springer, New York.
- EYRE-WALKER, A., M. WOOLFIT and T. PHELPS, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* **173**: 891–900.
- GALTIER, N., and L. DURET, 2007 Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* **23**: 273–277.
- GALTIER, N., G. PIGANEAU, D. MOUCHIROUD and L. DURET, 2001 GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* **159**: 907–911.
- GALTIER, N., E. BAZIN and N. BIERNE, 2006 GC-biased segregation of noncoding polymorphisms in *Drosophila*. *Genetics* **172**: 221–228.
- GALTIER, N., L. DURET, S. GLÉMIN and V. RANWEZ, 2009 GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* **25**: 1–5.
- GLÉMIN, S., 2003 How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* **57**: 2678–2687.
- GLÉMIN, S., 2007 Mating systems and the efficacy of selection at the molecular level. *Genetics* **177**: 905–916.
- GLÉMIN, S., J. RONFORT and T. BATAILLON, 2003 Patterns of inbreeding depression and architecture of the load in subdivided populations. *Genetics* **165**: 2193–2212.
- GLÉMIN, S., E. BAZIN and D. CHARLESWORTH, 2006 Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proc. Biol. Sci.* **273**: 3011–3019.
- GUTZ, H., and J. F. LESLIE, 1976 Gene conversion: a hitherto overlooked parameter in population genetics. *Genetics* **83**: 861–866.
- HALDANE, J. B. S., 1937 The effect of variation on fitness. *Am. Nat.* **71**: 337–349.
- HAUDRY, A., A. CENCI, C. GUILHAUMON, E. PAUX, S. POIRIER *et al.*, 2008 Mating system and recombination affect molecular evolution in four Triticeae species. *Genet. Res.* **90**: 97–109.
- INTERNATIONAL CHICKEN GENOME SEQUENCING CONSORTIUM, 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695–716.
- JANCEK, S., S. GOURBIERE, H. MOREAU and G. PIGANEAU, 2008 Clues about the genetic basis of adaptation emerge from comparing the proteomes of two *Ostreococcus* ecotypes (Chlorophyta, Prasinophyceae). *Mol. Biol. Evol.* **25**: 2293–2300.
- KEIGHTLEY, P., and A. EYRE-WALKER, 2000 Deleterious mutations and the evolution of sex. *Science* **290**: 331–333.
- KEIGHTLEY, P. D., M. J. LERCHER and A. EYRE-WALKER, 2005 Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol.* **3**: e42.
- KEIGHTLEY, P. D., M. J. LERCHER and A. EYRE-WALKER, 2006 Understanding the degradation of hominid gene control. *PLoS Comput. Biol.* **2**: e19 (author reply: e26).
- KIMURA, M., 1962 On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713–719.
- KIMURA, M., and T. OHTA, 1971 *Theoretical Aspects of Population Genetics*. Princeton University Press, Princeton, NJ.
- KIMURA, M., and T. OHTA, 1973 The age of a neutral mutant persisting in a finite population. *Genetics* **75**: 199–212.
- KIMURA, M., T. MARUYAMA and J. F. CROW, 1963 The mutation load in small populations. *Genetics* **48**: 1303–1312.
- KURAKU, S., J. ISHIJIMA, C. NISHIDA-UMEHARA, K. AGATA, S. KURATANI *et al.*, 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.
- LAMB, B. C., and S. HELMI, 1982 The extent to which gene conversion can change allele frequencies in populations. *Genet. Res.* **39**: 199–217.
- LANDE, R., 1994 Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**: 1460–1469.
- LANDE, R., 1998 Risk of population extinction from fixation of deleterious and reverse mutations. *Genetica* **102/103**: 21–27.
- LANDE, R., and D. W. SCHEMSKE, 1985 The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* **39**: 24–40.
- LYNCH, M., 2007 *The Origin of Genome Architecture*. Sinauer Associates, Sunderland, MA.
- LYNCH, M., J. CONERY and R. BURGER, 1995 Mutation accumulation and the extinction of small populations. *Am. Nat.* **146**: 489–518.
- MANCERA, E., R. BOURGON, A. BROZZI, W. HUBER and L. M. STEINMETZ, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**: 479–485.
- MARAIS, G., 2003 Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* **19**: 330–338.
- MARAIS, G., D. MOUCHIROUD and L. DURET, 2001 Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl. Acad. Sci. USA* **98**: 5688–5692.

- MARAIS, G., D. MOUCHIROUD and L. DURET, 2003 Neutral effect of recombination on base composition in *Drosophila*. *Genet. Res.* **81**: 79–87.
- MARAIS, G., B. CHARLESWORTH and S. I. WRIGHT, 2004 Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. *Genome Biol.* **5**: R45.
- MARSOLIER-KERGOAT, M. C., and E. YERAMIAN, 2009 GC content and recombination: reassessing the causal effects for the *Saccharomyces cerevisiae* genome. *Genetics* **183**: 31–38.
- MCVEAN, G. A., S. R. MYERS, S. HUNT, P. DELOUKAS, D. R. BENTLEY *et al.*, 2004 The fine-scale structure of recombination rate variation in the human genome. *Science* **304**: 581–584.
- MEUNIER, J., and L. DURET, 2004 Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.* **21**: 984–990.
- MYERS, S., L. BOTTOLO, C. FREEMAN, G. MCVEAN and P. DONNELLY, 2005 A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**: 321–324.
- NAGYLAKI, T., 1983a Evolution of a finite population under gene conversion. *Proc. Natl. Acad. Sci. USA* **80**: 6278–6281.
- NAGYLAKI, T., 1983b Evolution of a large population under gene conversion. *Proc. Natl. Acad. Sci. USA* **80**: 5941–5945.
- NIELSEN, R., 2005 Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**: 197–218.
- NISHANT, K. T., and M. R. S. RAO, 2005 Molecular features of meiotic recombination hotspots. *BioEssays* **28**: 45–56.
- NORDBORG, M., and P. DONNELLY, 1997 The coalescent process with selfing. *Genetics* **146**: 1185–1195.
- NORDBORG, M., B. CHARLESWORTH and D. CHARLESWORTH, 1996 The effect of recombination on background selection. *Genet. Res.* **67**: 159–174.
- OHTA, T., and C. C. COCKERHAM, 1974 Detrimental genes with partial selfing and effects on a neutral locus. *Genet. Res.* **23**: 191–200.
- POLLAK, E., 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* **117**: 353–360.
- POLLARD, K. S., S. R. SALAMA, B. KING, A. D. KERN, T. DRESZER *et al.*, 2006a Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**: e168.
- POLLARD, K. S., S. R. SALAMA, N. LAMBERT, M. A. LAMBOT, S. COPPENS *et al.*, 2006b An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**: 167–172.
- PRABHAKAR, S., J. P. NOONAN, S. PAABO and E. M. RUBIN, 2006 Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**: 786.
- PRABHAKAR, S., A. VISEL, J. A. AKIYAMA, M. SHOUKRY, K. D. LEWIS *et al.*, 2008 Human-specific gain of function in a developmental enhancer. *Science* **321**: 1346–1350.
- ROCKMAN, M. V., and L. KRUGLYAK, 2009 Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* **5**: e1000419.
- SPENCER, C. C., P. DELOUKAS, S. HUNT, J. MULLIKIN, S. MYERS *et al.*, 2006 The influence of recombination on human genetic diversity. *PLoS Genet.* **2**: e148.
- TAKAHATA, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc. Natl. Acad. Sci. USA* **87**: 2419–2423.
- TAKEBAYASHI, N., and P. L. MORRELL, 2001 Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *Am. J. Bot.* **88**: 1143–1150.
- UYENOYAMA, M. K., and D. M. WALLER, 1991 Coevolution of self-fertilization and inbreeding depression. II. Symmetric overdominance in viability. *Theor. Popul. Biol.* **40**: 47–77.
- WANG, L. J., and M. J. ROOSSINCK, 2006 Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. *Plant Mol. Biol.* **61**: 699–710.
- WEBSTER, M. T., E. AXELSSON and H. ELLEGREN, 2006 Strong regional biases in nucleotide substitution in the chicken genome. *Mol. Biol. Evol.* **23**: 1203–1216.
- WHITLOCK, M. C., P. K. INGVARSSON and T. HATFIELD, 2000 Local drift load and the heterosis of interconnected populations. *Heredity* **84**: 452–457.
- WOLFRAM, S., 1996 *The Mathematica Book*. Cambridge University Press, Cambridge, UK.
- WRIGHT, S., 1937 The distribution of gene frequencies in populations. *Proc. Natl. Acad. Sci. USA* **23**: 307–320.
- WRIGHT, S. I., G. IORGOVAN, S. MISRA and M. MOKHTARI, 2007 Neutral evolution of synonymous base composition in the Brassicaceae. *J. Mol. Evol.* **64**: 136–141.
- ZIEHE, M., and J. H. ROBERDS, 1989 Inbreeding depression due to overdominance in partially self-fertilizing plant populations. *Genetics* **121**: 861–868.

Communicating editor: S. G. HAMISH

APPENDIX A: GENERAL DERIVATIONS

Analysis of the deterministic model: Equation 1 gives the change in allele frequency due to mutation, selection, and gBGC, as follows:

$$\Delta x = u(1 - x) - \lambda ux + b(1 - F)x(1 - x) - sx(1 - x)(h + F - hF + (1 - F)(1 - 2h)x) \quad (\text{A1a})$$

if *S* is deleterious and

$$\Delta x = \lambda u(1 - x) - ux - b(1 - F)x(1 - x) - sx(1 - x)(h + F - hF + (1 - F)(1 - 2h)x) \quad (\text{A1b})$$

if *W* is deleterious. If *W* is deleterious, both selection and gBGC maintain the deleterious alleles at low frequency, so that $x \ll 1$. Linearizing Δx around 0 and neglecting back mutation gives $\Delta x \approx \lambda u - b(1 - F)x - sx(h + F - hF)$, which leads to (3). If *S* is the deleterious allele, this leads to three selection regimes. If gBGC is weak compared to selection ($b \ll s$), $x \ll 1$, and the linear approximation gives $\Delta x \approx u + b(1 - F)x - sx(h + F - hF)$, which leads to (4a). If gBGC is strong ($b \gg s$), x is very close to 1. Linearizing Δx around 1 and neglecting direct mutation gives $\Delta x \approx \lambda u + b(1 - F)(1 - x) - s(1 - x)(1 - h + hF)$, which leads to (4b). Finally, if gBGC and selection are of similar intensity, the two forces can maintain the deleterious allele at intermediate frequency and mutations can be neglected, $\Delta x \approx x(1 - x)(b(1 - F) - s(h + F - hF + (1 - F)(1 - 2h)x))$, which leads to (4c). In the special case of fully recessive alleles in panmixia ($F = 0$), using $x \ll 1$, Δx can be approximated by $\Delta x \approx u + \lambda ux + bx - sx^2$, which leads to (4d).

Analysis of the deterministic model: In finite populations, the WRIGHT (1937) stationary distribution is given by

$$\Phi(x) = \frac{K}{V_x} e^{\int 2(M_x/V_x)}, \quad (\text{A2})$$

where $M_x = \Delta_x$, given by Equation 1, and $V_x = x(1-x)/2N_e$ are the infinitesimal mean and variance of changes in allele frequency. K is an integration constant, ensuring that $\int_0^1 \Phi(x) dx = 1$ and $N_e = \alpha(N/(1+F))$. Integration of (A2) leads to (6a) and (6b).

To compute the probability of fixation, let us note

$$G(x) = e^{-\int (2M_x/V_x)}, \quad (\text{A3})$$

where $M_x = \Delta_{x_{\text{conversion}}} + \Delta_{x_{\text{selection}}}$ and V_x is the same as in (A2), and the probability of fixation starting from frequency p_0 is

$$P_{\text{fix}}(p) = \frac{\int_0^{p_0} G(x) dx}{\int_0^1 G(x) dx} \quad (\text{A4})$$

(KIMURA 1962).

For a new mutation ($p_0 = 1/2N$) under genic selection in panmixia, (A4) simply reduces to (7a) and (7b). General analytical expressions for P_{fix} can be obtained but they are formidable and not given here.

The mean sojourn and fixation times of a new mutation can be obtained using classical diffusion results (e.g., EWENS 2004),

$$\bar{T}_{\text{soj}} = (1 - P_{\text{fix}}) \int_0^{p_0} P_{\text{fix}}(p)(1 + P_{\text{fix}}(p))\varphi(p) dp + P_{\text{fix}} \int_{p_0}^1 (1 - P_{\text{fix}}(p))(2 - P_{\text{fix}}(p))\varphi(p) dp \quad (\text{A5})$$

and

$$\bar{T}_{\text{fix}} = \frac{1 - P_{\text{fix}}}{P_{\text{fix}}} \int_0^{p_0} P_{\text{fix}}(p)^2 \varphi(p) dp + \int_{p_0}^1 P_{\text{fix}}(p)(1 - P_{\text{fix}}(p))\varphi(p) dp, \quad (\text{A6})$$

where $\varphi(p) = \int_0^1 G(\xi) d\xi / V_p G(p)$ with G defined in (A3), and $p_0 = 1/2N$.

The mean time until the mutant allele reaches the deterministic equilibrium frequency, x_{eq} , for the first time is given by the KIMURA and OHTA (1973) expression, which gives the mean time to reach a given frequency for the first time, starting from frequency $p_0 < x_{\text{eq}}$,

$$\bar{T}_{\text{reach}} = \frac{1 - v(x_{\text{eq}}, p_0)}{v(x_{\text{eq}}, p_0)} \int_0^{p_0} v(x_{\text{eq}}, p)^2 \psi(p) dp + \int_{p_0}^1 v(x_{\text{eq}}, p)(1 - v(x_{\text{eq}}, p))\psi(p) dp, \quad (\text{A7})$$

where

$$v(x_{\text{eq}}, p) = \frac{\int_0^p G(x) dx}{\int_0^{x_{\text{eq}}} G(x) dx}, \quad \psi(x_{\text{eq}}, p) = \frac{\int_0^{x_{\text{eq}}} G(\xi) d\xi}{V_p G(p)}, \quad \text{and} \quad p_0 = 1/2N.$$

For fully recessive mutations in panmixia, the equilibrium frequency is well approximated by $x_{\text{eq}} = b/(b+s)$ (Equation 4c). Under the overdominance-like regime, if the allele is not initially lost, it can persist for a long time (e.g., TAKAHATA 1990). To evaluate this conditional persistence, I also computed the mean sojourn time of an allele starting from frequency x_{eq} , using $p_0 = x_{\text{eq}}$ in (A5).

In the general case, these expressions have no analytical expression and numerical integrations were performed (see APPENDIX B).

APPENDIX B: NUMERICAL INTEGRATIONS

Stationary distribution and time properties: Because $\Phi(x)$ given by Equations 5a and Equations 5b can diverge in 0, 1, or both, I adapted KIMURA *et al.*'s (1963) quadrature method to facilitate the numerical integration of $E_\Phi[x^i]$. The principle is given for S deleterious alleles.

Let $q_i(x) = x^{4(\alpha N/(1+F))u-1+i}$ and $r(x) = (1-x)^{4(\alpha N/(1+F))\lambda u-1} e^{-2(\alpha N/(1+F))(s(1+F)-2b(1-F))}$, which verifies

$$x^i \Phi(x) \xrightarrow{x \rightarrow 0} Kq_i(x) \quad \text{and} \quad x^i \Phi(x) \xrightarrow{x \rightarrow 0} Kr(x).$$

Let

$$Q_i = \int_0^1 q_i(x) dx = \frac{1}{i + 4(\alpha N / (1 + F))u}$$

and

$$R = \int_0^1 r(x) dx = \frac{e^{2(\alpha N / (1 + F))(2b(1 - F) - (1 + F)s)}}{4(\alpha N / (1 + F))\lambda u}.$$

Noting $\phi(x) = \Phi(x)/K$, we thus have

$$\frac{1}{K} = \int_0^1 (\phi(x) - q_0(x) - r(x)) dx + Q_0 + R$$

and

$$E_\Phi[x^i] = K \int_0^1 (\phi(x) - q_i(x) - r(x)) dx + Q_i + R.$$

Due to this manipulation, the functions that are numerically integrated do not diverge. Numerical integrations were done using the NIntegrate function of the Mathematica software (WOLFRAM 1996), setting the maximum recursion at 50, working precision at 15, and other options at default values.

The other integrals converge so I performed simple numerical integrations using the NIntegrate function of the Mathematica software (WOLFRAM 1996), setting maximum recursion at 50, working precision at 15, and other options at default values.

Total genomic load: To compute the load and on the whole-genome scale, I integrated Equation 11 over a distribution of mutational effects. I assumed a gamma distribution of selection coefficients, $\psi(s)$, with mean $s_{\text{mean}} = 0.0325$ and shape parameter $= 0.23$, according to EYRE-WALKER *et al.* (2006). Since the load weakly depends on the dominance levels, I assumed $h = \frac{1}{2}$. For simplicity, I assumed that the fitness is multiplicative, such that

$$L_{\text{genomic}} = 1 - \prod_{i=1}^n (1 - L_i) \approx 1 - e^{-n \int L\psi(s) ds},$$

where L_i is the load contributed by the locus i , and n is the number of loci contributing to the load. I used $u = 10^{-8}$ and $n = 10^6$ so that the genomic deleterious mutation rate is $U = 10^{-2}$. This is much lower than current estimates in humans ($U > 1$, KEIGHTLEY and EYRE-WALKER 2000; KEIGHTLEY *et al.* 2005, 2006). However, using such values leads to unbearable loads under the multiplicative fitness model. I performed numerical integration using the quadrature method (see above) and the function NIntegrate of Mathematica software, setting maximum recursion at 50 and working precision at 15.

APPENDIX C: LOAD APPROXIMATIONS

Genic selection in panmixia: For genic selection in panmixia, an interesting approximation can be obtained for the load as follows. Consider that the deleterious allele can be either fixed with probability p_F , causing the load s , or segregating at deterministic frequency with probability $1 - p_F$, causing the load L_{det} . The probability of being fixed for the deleterious allele, p_F , is approximately given by

$$p_F \approx \frac{uP_{\text{fix}}(s, b)}{uP_{\text{fix}}(s, b) + vP_{\text{fix}}(-s, -b)} \quad (\text{C1})$$

for S deleterious alleles and

$$p_F \approx \frac{vP_{\text{fix}}(s, -b)}{uP_{\text{fix}}(-s, b) + vP_{\text{fix}}(s, -b)} \quad (\text{C2})$$

for W deleterious alleles (*e.g.*, BULMER 1991; LANDE 1998).

Under genic selection and panmixia, this reduces to

$$p_F \approx \frac{1}{1 + \lambda e^{2N_e(s-2b)}} \quad (C3)$$

if S is the deleterious allele and

$$p_F \approx \frac{1}{1 + e^{2N_e(s+2b)}/\lambda} \quad (C4)$$

if W is the deleterious allele (see also KIMURA *et al.* 1963; BULMER 1991; LANDE 1998).

This leads to Equations 12a and Equations 12b.

Maximum load: The deleterious effect of mutations that maximizes the load can be obtained by solving in s : $\partial L/\partial s = 0$. Approximate solutions can be obtained through the following rationale (for the case without gBGC, see also KIMURA *et al.* 1963). Mutations that cause the highest load are the most deleterious ones among those that can reach fixation. Fixation can be reached because of drift or gBGC. We can thus simply maximize the first part of Equations 12, that is, $s p_F$. The selection coefficients that maximize the load, $s_{\max}$, are thus solutions in s of

$$\frac{\lambda(\lambda + (1 - 2Ns)e^{2N(2b+s)})}{(\lambda + e^{2N(2b+s)})^2} = 0 \quad (C5)$$

for W deleterious alleles and

$$\frac{1 + \lambda(1 - 2Ns)e^{2N(s-2b)}}{(1 + \lambda e^{2N(s-2b)})^2} = 0 \quad (C6)$$

for S deleterious alleles.

This leads to

$$s_{\max} = \frac{1}{2N} (1 + \text{PL}(\lambda e^{-(1+4Nb)})) \quad (C7)$$

if W is the deleterious allele and

$$s_{\max} = \frac{1}{2N} \left(1 + \text{PL} \left(\frac{e^{-(1-4Nb)}}{\lambda} \right) \right) \quad (C8)$$

if S is the deleterious allele, where PL is the product log function (ABRAMOWITZ and STEGUN 1970).

For the W deleterious allele, $1/2N < s_{\max} < (1/2N)(1 + \text{PL}(\lambda e^{-1}))$, which is $\sim 1/2N < s_{\max} < 1/N$ for realistic λ values. For the S deleterious allele, $(1/2N)(1 + \text{PL}(e^{-1}/\lambda)) < s_{\max} < (1/2N)(\text{PL}(e^{4Nb}/\lambda))$. For realistic λ -values, the minimum value is composed between $1/2N$ and $1/N$ and the maximum one between b and $2b$ if $Nb \gg 1$. So we have $\sim \min(1/2N, b) < s_{\max} < \max(1/N, 2b)$.

Minimum load: Similarly, one can compute the gBGC level that minimizes the total load by solving in b : $\partial L/\partial b = 0$. Here, I compute separately the gBGC level that minimizes the segregating load and the fixation load. gBGC weakly decreases the load due to W deleterious mutations. To minimize the total segregating load, the gBGC level must weakly increase the load due to S deleterious alleles, so $b < hs$. Using Equation 10a, we thus need to minimize $q(2uhs/(hs - b)) + (1 - q)(2\lambda uhs/(hs + b))$. The solution is

$$b_{\min}^{\text{seg}} = hs \frac{(q(\lambda - 1) - \lambda + 2\sqrt{q(1 - q)\lambda})}{q - (1 - q)\lambda}, \quad (C9)$$

which reduces to

$$b_{\min}^{\text{seg}} = hs \left(1 - \frac{2}{1 + \sqrt{\lambda}} \right) \quad (C10)$$

for $q = \frac{1}{2}$.

FIGURE D1.— $N = 10,000$; $b = 0$, $b = 2 \times 10^{-3}$, $b = 2 \times 10^{-4}$, $b = 5 \times 10^{-5}$ (from left to right).

FIGURE D2.— $b = 2 \times 10^{-4}$; $s = 10^{-2}$, $s = 10^{-3}$, $s = 10^{-4}$, $s = 10^{-5}$ (from top left to bottom left).

As expected, gBGC can minimize the load only when $\lambda > 1$, that is, when the mutation is biased toward AT. The gBGC level that minimizes the drift load is given by the value $b_{\min}^{\text{fix}}$ that minimizes

$$q \frac{s}{1 + \lambda e^{2N(s-2b)}} + (1-q) \frac{s}{1 + e^{2N(s+2b)}/\lambda}.$$

The solution is

$$b_{\min}^{\text{fix}} = \frac{\ln(\lambda(e^{2Ns}(1-2q) + (e^{4Ns} - 1)\sqrt{q(1-q)})/(q(e^{4Ns} + 1) - 1))}{4N}, \quad (\text{C11})$$

which reduces to

$$b_{\min}^{\text{fix}} = \frac{\ln(\lambda)}{4N} \quad (\text{C12})$$

for $q = \frac{1}{2}$. Interestingly, (C12) is independent of s . In most cases, $b_{\min}^{\text{fix}} < b_{\min}^{\text{seg}}$; $b_{\min}^{\text{fix}}$ is thus a very good approximation of the gBGC level that minimizes the total load, even when assuming a distribution of selection coefficients (APPENDIX D).

Finally, we can compute gBGC levels above which it increases the fixation load by solving in b ,

$$q \frac{s}{1 + \lambda e^{2N(s-2b)}} + (1-q) \frac{s}{1 + e^{2N(s+2b)}/\lambda} = 0,$$

which leads to

$$b_{\text{lim}} = \frac{\ln(\lambda)}{2N}. \quad (\text{C14})$$

In most cases, we also have $b_{\text{lim}} < b_{\min}^{\text{seg}}$, so gBGC levels that decrease the segregating load should increase the fixation load.

APPENDIX D: NUMERICAL VALIDATION OF EQUATIONS 12 AND 14

Equations 12: Equations 12 are approximations for the exact Equation 10 that gives the load due to codominant mutations ($h = \frac{1}{2}$) in panmixia. Figure D1 and Figure D2 show that, for S deleterious alleles, (12a) is a very accurate approximation for (11) for a wide range of parameters. This is also true for W deleterious alleles (not shown):

- (a) $N = 10,000$; $b = 0$, $b = 2 \times 10^{-3}$, $b = 2 \times 10^{-4}$, $b = 5 \times 10^{-5}$ (from left to right).
- (b) $b = 2 \times 10^{-4}$; $s = 10^{-2}$, $s = 10^{-3}$, $s = 10^{-4}$, $s = 10^{-5}$ (from top left to bottom left).

Equation 14: The total genomic load is computed as explained in APPENDIX B. Then the value of b that minimizes the load was numerically computed through a simple iterative algorithm. Numerical results were compared to Equation 14. Results are given ($\times 10^{-3}$) for various mutational biases (λ), population sizes (N), and shape of the gamma distribution of deleterious mutations (β). (See Table 6).

TABLE 6
gBGC levels minimizing the total genomic load

	Population size	Equation 14	Numeric $\beta = 0.01$	Numeric $\beta = 0.23$	Numeric $\beta = 1$
$\lambda = 2$	$N = 1,000$	0.173	0.174	0.174	0.174
	$N = 10,000$	0.0173	0.0174	0.0174	0.0174
	$N = 100,000$	0.00173	0.00178	0.00179	0.00181
$\lambda = 5$	$N = 1,000$	0.402	0.403	0.403	0.403
	$N = 10,000$	0.0402	0.0404	0.0406	0.0410
	$N = 100,000$	0.00402	0.00411	0.00412	0.00434
$\lambda = 10$	$N = 1,000$	0.576	0.577	0.577	0.577
	$N = 10,000$	0.0576	0.0579	0.0579	0.0581
	$N = 100,000$	0.00576	0.00596	0.00597	0.00646