Evolvability of an Optimal Recombination Rate

Alexander E. Lobkovsky[1], Yuri I. Wolf[1], and Eugene V. Koonin[1,*]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland

*Corresponding author: E-mail: koonin@ncbi.nlm.nih.gov.

Abstract

Evolution and maintenance of genetic recombination and its relation to the mutational process is a long-standing, fundamental problem in evolutionary biology that is linked to the general problem of evolution of evolvability. We explored a stochastic model of the evolution of recombination using additive fitness and infinite allele assumptions but no assumptions on the sign or magnitude of the epistasis and the distribution of mutation effects. In this model, fluctuating negative epistasis and predominantly deleterious mutations arise naturally as a consequence of the additive fitness and a reservoir from which new alleles arrive with a fixed distribution of fitness effects. Analysis of the model revealed a nonmonotonic effect of recombination intensity on fitness, with an optimal recombination rate value which maximized fitness in steady state. The optimal recombination rate depended on the mutation rate and was evolvable, that is, subject to selection. The predictions of the model were compatible with the observations on the dependence between genome rearrangement rate and gene flux in microbial genomes.

Key words: recombination, allele replacement, evolvability, gene flux, genome, rearrangement.

Introduction

Genetic recombination is a universal biological mechanism. All cellular life forms encode enzymatic machinery for recombination, and in sexually reproducing organisms multiple recombination events occur during each cycle of reproduction, in the course of meiosis (Friedberg et al. 2005). The ubiquity of recombination poses a theoretical challenge given its cost that results from breaking up combinations of alleles favored by selection (Fisher 1930; Bodmer and Parsons 1962). Although the prevalence of recombination might have a proximal or mechanistic explanation, such as its role in DNA repair, many studies have focused on evolutionary explanations under which recombination confers a fitness advantage that under certain conditions could offset the cost (Barton and Charlesworth 1998). In 1967, Nei proposed the general modifier allele framework for the study of the fitness effect of recombination and showed that selection for recombination was possible only in the presence of interaction between genes (epistasis) (Nei 1967). Recombination can provide at least two distinct evolutionary benefits (Barton 1995). The short-term benefit is in breaking apart less-fit combinations of genes whereas the long-term benefit is in increasing the fitness variance of the offspring thereby increasing the efficacy of selection. In either case, selection on the recombination modifier operates through its association with directly selected loci. Whether or not recombination can provide either the short- or the long-term benefit, or both, depends on the character of the linkage disequilibria that exist in the absence of recombination and the nature of epistasis (Kondrashov 1993).

When the fitness landscape is constant in time, selection favors linkage disequilibrium of the same sign as the prevailing epistasis of the landscape (Felsenstein 1965; Eshel and Feldman 1970). Because pairs of alleles that have higher combined fitness occur at a higher frequency, the short-term effect of recombination which breaks up these beneficial associations is deleterious. Nevertheless, recombination can still be beneficial in the long term as long as epistasis is negative and the short-term recombination load is not prohibitively high (Feldman et al. 1980; Barton 1995; Otto and Feldman 1997).

Under the red-queen hypothesis, fluctuating epistasis results in a mismatch between the current frequent allele combinations and those favored by selection (Sturtevant and Mather 1938; Charlesworth 1976; Maynard Smith 1976; Barton 1995). Recombination can then be beneficial in the short term by breaking up unfavorable allele combinations. However, both selection and the frequency and magnitude of the epistasis fluctuations need to be strong for this mechanism to operate resulting in a restricted parameter range of validity (Charlesworth 1976; Barton 1995; Peters and Lively 1999; Gandon and Otto 2007).

Fluctuations resulting from a finite population size lead to a persistent negative linkage disequilibrium (Hill and Robertson 1966), so that recombination is favored for its long-term benefit because it increases genetic variation and restores high fitness allele combinations that have been lost through drift (Felsenstein and Yokoyama 1976; Otto and Barton 2001; Iles et al. 2003; Keightley and Otto 2006; Martin et al. 2006). This mechanism requires intermediate population sizes to operate although spatial structure or a large number of selected loci both increase the upper population size limit at which recombination remains beneficial (Iles et al. 2003; Martin et al. 2006). Spatial structure affects recombination even in the absence of drift. The spatial heterogeneity of selection coupled with migration generates linkage disequilibrium dependent on the sign of the epistasis. Therefore, recombination can be either deleterious or beneficial (Pylkov et al. 1998; Martin et al. 2006; Roze 2009; Hartfield et al. 2012; Takeuchi et al. 2014). A model with recurrent deleterious mutations and complete linkage between loci exhibited selection on a sex modifier sufficiently strong to overcome a 2-fold cost, but only when sex (recombination) was rare and the modifier led to modest increases in its frequency (Keightley and Otto 2006; Hartfield et al. 2012).

Theoretical models usually predict a monotonic fitness effect of recombination and strive to identify the conditions under which recombination is selectable. However, the existence of a nontrivial equilibrium level of recombination maintained by selection has been demonstrated in a variety of contexts of variable complexity. In particular, a nonmonotonic effect of recombination on fitness has been reported in a model with a large number of loci under selection (Iles et al. 2003). An optimal recombination rate also has been found to exist in a model with cyclical epistasis (Gandon and Otto 2007).

The rate of homologous recombination is difficult to measure directly. In contrast, the genomic rearrangement rate is readily measurable (Novichkov, Wolf, et al. 2009). Because the dominant molecular mechanism for rearrangement is homologous recombination (Darmon and Leach 2014), we use the rearrangement rate as a proxy for the homologous recombination rate.

Measurements presented here reveal an upper bound on the rearrangement rate that grows sublinearly with the rate of single gene replacements which we denote the gene flux. This observation is consistent with a Wright–Fisher model with negative epistasis that exhibits an optimal, gene flux-dependent recombination rate maintained by selection.

Materials and Methods

The Alignable Tight Genomic Clusters (ATGC) database (Novichkov, Ratnere, et al. 2009; Puigbo et al. 2014) assembles clusters of orthologous genes (COGs) (Tatusov et al. 1997; Galperin et al. 2015) for each group of genomes and maps each protein to a unique cluster. The protein–COG mapping was used to construct gapped local gene-by-gene alignments (Wolf et al. 2001) that typically covered over 90% of the genome. The genome alignments were constructed by assigning a score of 1 to a pair of genes in the same COG and 0 otherwise. Given penalties for a mismatch, gap opening and gap extension, matching segments are merged if the total score exceeds the sum of segment scores prior to merging. All genome alignments were computed for the three penalties set to 0.1 (see Wolf et al. [2001] for details).

When the genome rearrangement rate is smaller than the ratio of the genome size and the typical size of a rearrangement, the number of aligned segments is linear in the number of rearrangement events; accordingly, the number of segments was used as a proxy for the number of rearrangement events. With the chosen scoring system, the resulting alignments can, in principle, span gaps or mismatched regions an order of magnitude longer than the matching segments. Therefore, the alignment is tolerant to most of the gene-scale disturbances and breaks only at the sites of major recombination events, such as chromosomal inversions or translocations into entirely different context.

Gene flux was defined as the sum of the number of gaps and mismatches in the aligned genomic segments divided by the sum of the lengths of the aligned segments. The number of rearrangements per-genome and the gene flux per gene estimates were normalized by the median number of synonymous nucleotide substitutions (dS) among genes that were conserved in a particular ATGC. This normalization transformed raw counts to rates measured relative to the rate of (nearly) neutral substitutions and allowed the comparison of gene flux and rearrangement rates between different ATGCs.

Importantly, although the measured genome rearrangement rate is expected to be linearly correlated with the homologous recombination rate, the coefficient of proportionality is unknown a priori. In addition, the gene flux, that is, the rate at which accumulated mutations result in a qualitative allele change is different from the underlying nucleotide substitution rate. It is therefore not possible to compute the traditional recombination–mutation ratio (r/m) (Guttman and Dykhuizen 1994).

Results

Dependency of Genome Rearrangement Rate on Gene Flux in Bacterial and Archaeal Genomes

We were prompted to investigate a mathematical model of the evolution of recombination by an empirical observation on bacterial and archaeal genomes. By comparing the genomes within multiple groups of closely related bacteria and archaea, we found that the upper bound of the rate of genome rearrangement scaled sublinearly with the rate of replacements of individual genes.

Comparative analysis of the growing number of diverse bacterial and archaeal genomes shows that gene flux, that is, the loss and gain of individual genes, and genome rearrangement are the primary modes of microevolution in prokaryotes (Koonin and Wolf 2008). Although the rates of loss and gain vary substantially, a typical gene family is lost and gained dozens of times during the period in which sequence divergence reaches a unit substitution per site (Puigbo et al. 2014). Typical genome rearrangement rates are not as well known but a long-term evolution experiment in *Escherichia coli* suggests that the rate could be as high as one event per 10,000 generations (Raeside et al. 2014). We performed systematic measurements of the gene flux and genome rearrangement intensities using the ATGC database of groups of closely related genomes of bacteria and archaea (Novichkov, Ratnere, et al. 2009) (see supplementary file S1, Supplementary Material online, and Materials and Methods for details; the genome alignments are available at ftp://ftp.ncbi.nlm.nih.gov/pub/wolf_suppl/recombination; last accessed December 22, 2015). We assume that individual genes are gained and lost through mechanisms that are distinct from those responsible for the translocation, insertion, deletion, and inversion of large (at the scale of tens of genes or more) genomic segments. Genomic rearrangement is caused largely by recombination events occurring within the same genome, such as genome segment inversions centered at the origin of replication (Eisen et al. 2000; Novichkov, Wolf, et al. 2009). In contrast, the gene flux involves replacement of individual alleles by novel alleles that originate from a large gene reservoir, the supergenome (Lobkovsky et al. 2014). Thus, the degree of segmentation of the whole-genome alignment is associated with the intensity of rearrangement whereas gene mismatches and gaps found in aligned segments are attributed to the gene flux.

Figure 1 shows the dependency of the genome rearrangement rate on the gene flux for all analyzed ATGCs. The rearrangement rate exhibits an upper bound which scales sublinearly with the gene flux. In addition, the variance of the rearrangement rate notably declines with the increasing gene flux. We sought an explanation for this dynamic link between the two rates that does not invoke a common mechanism or a common control on the processes of rearrangement and flux, but possibly could be maintained by selection.

Modeling the Evolution of Recombination

If the effect of recombination on fitness is monotonic and recombination rate can evolve, selection will drive the rate to zero (if recombination is deleterious) or to infinity (if recombination is beneficial). Thus, models that predict a monotonic dependency of fitness on recombination intensity cannot explain the observed upper bound on the recombination rate (fig. 1) without invoking an additional mechanism that would limit the rate from above in a way that is dependent on the

gene flux. Considering the apparent implausibility of such mechanisms, we sought to construct a model that would naturally exhibit a nonmonotonic effect of recombination. If there exists an optimal recombination rate which maximizes fitness, selection can maintain the rate in the neighborhood of the optimum. The empirical observations could be explained if the optimal recombination rate depended sublinearly on the gene flux. Here we construct and analyze a stochastic population model which exhibits a gene flux-dependent, evolvable optimal recombination rate.

Fundamentally, recombination acts to reduce linkage disequilibrium by breaking associations between genes. If gene association is maintained by selection, recombination is deleterious and its effect is denoted as recombination load. Recombination will have a nonmonotonic effect on fitness if, at high recombination rates, the fitness effect is negative due to the disruption of beneficial gene associations, whereas at low recombination rates, the effect is positive owing to the segregation of deleterious alleles.

The following population model possesses the ingredients that lead to a nonmonotonic effect of recombination on steady-state fitness. The population of a fixed size N is composed of genomes whose fitness is the sum of individual contributions of alleles at a fixed number M of loci (additive fitness). Evolution is modeled through a Wright–Fisher process with discrete, nonoverlapping generations whereby the new generation is formed by selecting with replacement N organisms with probabilities proportional to their finesses. After the sampling, each gene of each organism is replaced by a novel
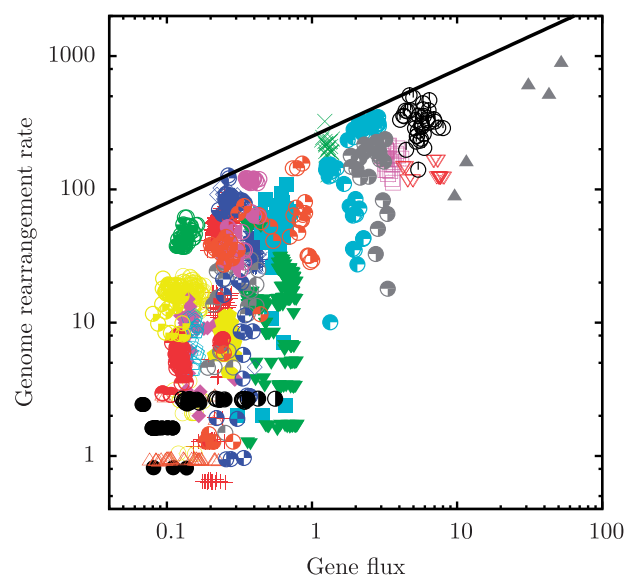
Fig. 1.—The dependency between the genome rearrangement rate and gene flux in bacteria and archaea. The solid line has a slope of 0.5. The data are presented for 27 ATGCs (each shown by a unique symbol) each containing at least ten microbial genomes (see supplementary file S1, Supplementary Material online).

allele with probability $1 - e^{-\mu}$. The contribution of the replacement alleles to the additive fitness is drawn from an exponential distribution with a unit mean, within an infinite allele pool. The results did change qualitatively when the distribution of fitness effects in the allele pool is uniform (supplementary figs. S1–S3, Supplementary Material online). After the gene replacement round, each organism replaces, with probability $1 - e^{-r}$, a genomic region of random uniformly distributed length starting from random position k (genomes are assumed to be circular) with the corresponding region of a random organism from the current generation.

Furthermore, replacing the uniformly distributed recombination region length by an exponentially distributed length did not alter the conclusions qualitatively (supplementary fig. S5, Supplementary Material online). The mean length of the recombining region was shown to have a weak effect on the steady-state fitness, with an optimum length that was an appreciable fraction of the genome size. Because our measurements of the genomic rearrangements revealed that the mean segment length was between 5% and 20% of the genome size, we proceed with the exploration of the model in which the length of the recombining region is uniformly distributed.

Simulations of the Wright–Fisher model were carried out until the steady-state criterion was satisfied. This criterion operated on the set of measurements of the mean fitness made every N generations. Simulations were terminated when the means among the middle third (by time) and the last third of these measurements were the same. To supplement the termination criterion, we fitted the time-dependent fitness to the expected logarithmic growth law with and without saturation and used the Akaike Information Criterion to determine whether saturation was supported by the data.

The infinite allele replacement mechanism causes negative linkage disequilibrium because a newly introduced allele occurs against a background where other loci are more likely to contain high frequency alleles. In addition, because selection drives the average fitness contribution significantly above unity, whereas the newly introduced alleles have a unit mean fitness, the effective fitness landscape exhibits negative epistasis (following the definition of Otto and Lenormand [2002]), in agreement with experimental evolution findings (Khan et al. 2011). Importantly, negative epistasis in our model is a natural consequence of the additive fitness assumption and does not emerge under the multiplicative fitness model. An extensive body of experimental evidence indicates that negative epistasis is the norm in evolution (Weinreich et al. 2005; Kouyos et al. 2007; Khan et al. 2011). The additive fitness assumption is a simple and natural way to construct a model with negative epistasis without having to introduce such epistasis explicitly.

The simulation results show that under this model, steady-state fitness has a pronounced peak as a function of the recombination rate r (fig. 2). The shape of the dependency as

well as the position and height of the peak depend on the gene flux. At low gene fluxes, high recombination rates are substantially deleterious, with steady-state fitness dropping below the level at zero recombination; in contrast, at high gene fluxes, the peak nearly degenerates into a plateau, that is, high recombination rates are almost as beneficial as the optimal value (fig. 2). The maximum benefit of recombination compared with the $r = 0$ case is observed at the intermediate values of the gene flux (fig. 3). Indeed, when the gene flux is so high that most alleles are replaced in each generation, the mean fitness contribution of genes in a genome is close to unity, and therefore epistasis is weak and the preservation of the association between the genes confers little or no fitness benefit. In contrast, when the gene flux is low (rare replacements), and drift is not too strong, selection can maintain linkage disequilibrium close to the optimal value. In both situations, reduction of disequilibrium caused by increased recombination does not increase the mean fitness.

In the model, the recombination rate r_{opt}, at which the maximum benefit is achieved, scales sublinearly with the gene flux (fig. 4). Qualitatively, this is the same dependency as was observed in the ATGC analysis through comparison of the measured genome rearrangement (recombination) rates and gene fluxes (fig. 1). A direct quantitative comparison of the model predictions to the empirically observed upper

$N = 200, M = 50$
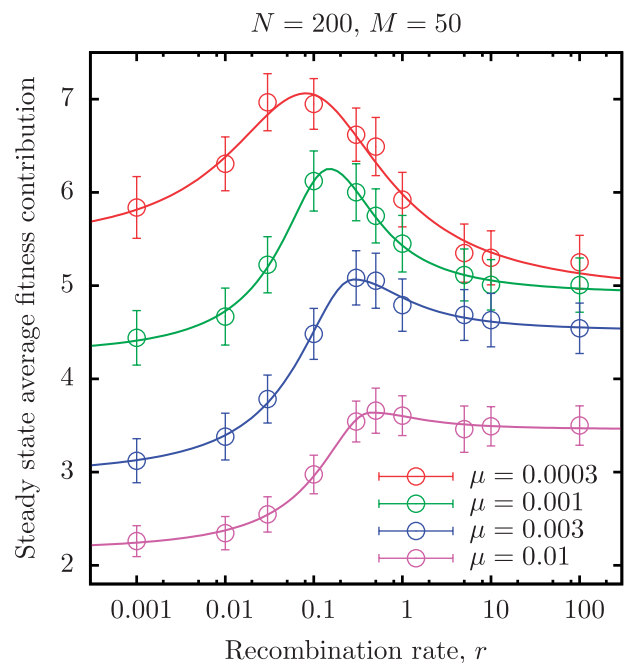
Fig. 2.—Steady state fitness of a model population as a function of the recombination rate. The simulation results are plotted for four gene flux values (color coded). Error bars display the standard deviation of the fitness from its mean in steady state.

bound on the genome rearrangement rate would require knowledge of the population size and generation time for each species. Neither of these values is readily available, and they are likely to exhibit large variability even within an ATGC.

The conclusion that the observed link between the genome rearrangement rate and gene flux results from the existence of an optimal recombination rate, which depends on the gene flux, implies that, when the recombination rate is allowed to evolve, selection can maintain its value near the optimum. This is not a foregone conclusion because, although the average fitness of a population in steady state is increased by recombination, the effect of recombination on the fitness of a genome can be negative because most low frequency alleles are deleterious.

We further tested the ability of selection to maintain the recombination rate close to the optimum by making the rate a heritable, mutable trait. In each generation, organisms have a probability ν of a mutation that changes their recombination rate by a small amount. If r is not subject to selection, the population average recombination rate performs a simple random walk. In contrast, if the trajectory of the population average recombination rate is consistent with diffusion in a potential well, one could conclude that selection maintains r close to an optimal value that corresponds to the bottom of the well. To test whether the trajectory $r(t)$ is consistent with simple diffusion, we plotted the autocorrelation function $\langle (r(t+\tau) - r(t))^2 \rangle$ (the averaging is done over time t) as a function of τ (fig. 5A). This quantity is expected to grow

linearly with time for simple diffusion, whereas our simulations clearly show saturation which is indicative of a steady state. The inset shows a typical trajectory $r(t)$ with the initial condition $r(0) = 0.001$. Figure 5B shows the probability distribution of the recombination rate in steady state. The distribution shows a peak at the optimal value of the recombination rate. The width of the peak decreases with increasing gene flux which is compatible with the empirical observations (fig. 1).

Discussion

Evolution and maintenance of recombination is a long standing, often vigorously debated theme in evolutionary biology (Fisher 1930; Muller 1964; Felsenstein 1974; Barton and Charlesworth 1998; Keightley and Otto 2006). This debate is linked to the more general and even more controversial problem of evolution of evolvability (Kirschner and Gerhart 1998; Draghi et al. 2010). The question whether evolvability mechanisms, of which recombination is arguably the most common one, are evolvable (subject to selection), has a long, chequered history. General considerations based on the popular belief that "evolution has no foresight" are often taken as an indication that such mechanisms, having no immediate effect on fitness, cannot be targets of selection (Partridge and Barton 2000; Chicurel 2001). Under this view, recombination would evolve as a by-product of essential processes of DNA repair. However, several modeling studies suggest that evolvability could be selectable (Earl and Deem 2004; Jones et al. 2007) or could evolve neutrally (Lehman and Stanley 2013), and a general mathematical solution to this problem has been proposed (Toussaint 2003).
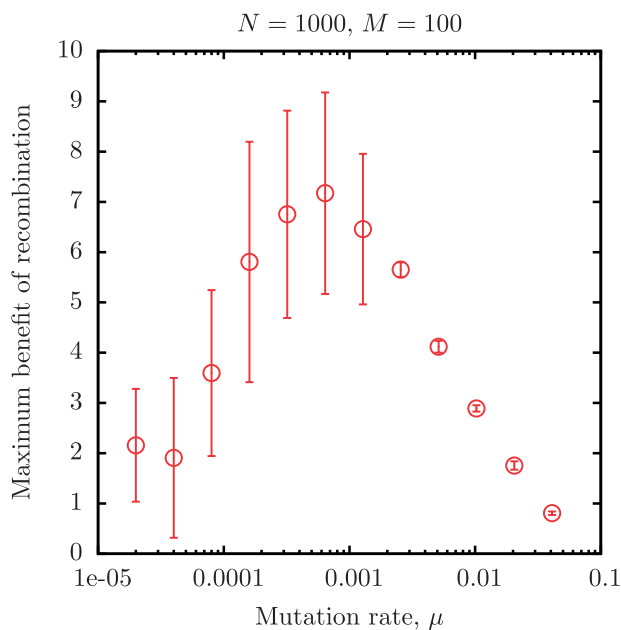
FIG. 3.—The maximum benefit of recombination versus the gene flux obtained by fitting the mean fitness as a function of the recombination rate.
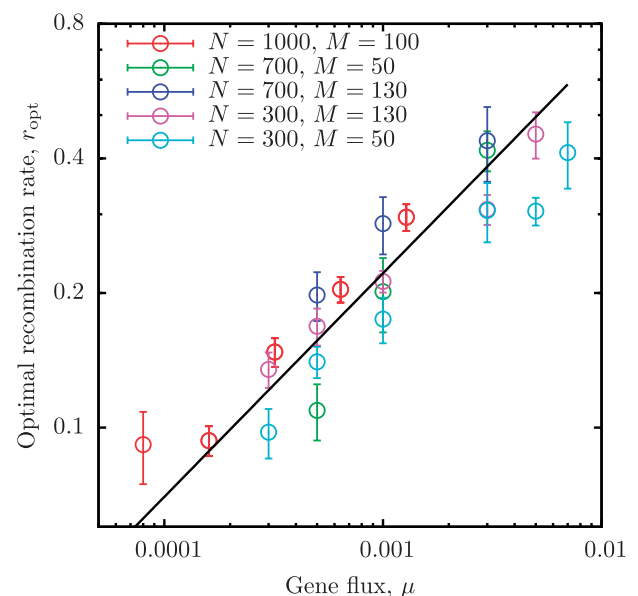
FIG. 4.—The optimal recombination rate in the model as a function of the gene flux.
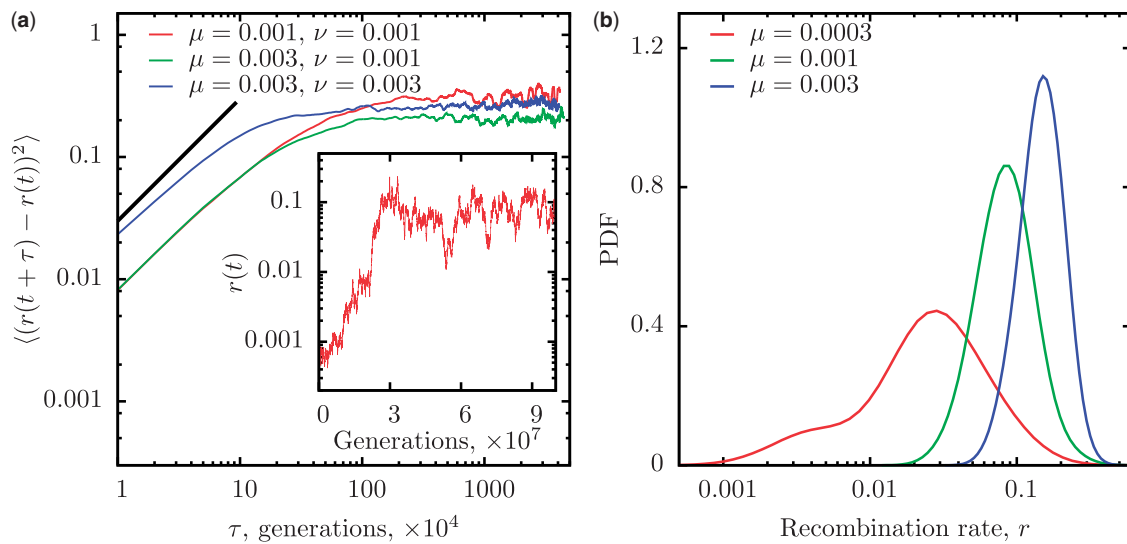
Fig. 5.—Evolution of the recombination rate in the model. (*a*) The main figure shows a plot of the autocorrelation function of the recombination rate against the number of generations for three combinations of gene flux values and the mutation rates of *r*. The inset shows a typical evolutionary trajectory of the recombination rate under the model. (*b*) Probability distribution function of the recombination rate produced by the model depending on the gene flux. Model parameters: $N = 10,000$; $M = 100$.

The fitness effect of recombination in stochastic population models depends on various factors, such as the structure and dynamics of the fitness landscape, mutation mechanisms and rates and the strength of the genetic drift, and can be either positive or negative (Christiansen et al. 1998; Otto and Lenormand 2002; Peng et al. 2004; Martin et al. 2006; Barton 2010; Rouzine and Coffin 2010; Weissman and Barton 2012; John and Jain 2015). A nonmonotonic fitness effect of recombination has been reported in a number of different scenarios (Iles et al. 2003; Keightley and Otto 2006; Gandon and Otto 2007).

The model described here has all the ingredients required to manifest a nonmonotonic fitness effect of recombination rate, and our evolutionary simulations indeed reveal the existence of optimal recombination rates that depend on the gene flux. Moreover, we explicitly show that, within the framework of this model, the optimal recombination rate is evolvable. Thus, in general terms, the results of the present work are compatible with the existence of selection for evolvability. Because the main focus of the previous studies of the fitness effect of recombination was on the emergence of recombination, the existence of an optimal recombination rate, although reported on several occasions, has not been thoroughly explored. Our aim was to investigate the dependence of the optimal recombination intensity on other parameters, such as allele replacement rate and population size, in an attempt to account for the empirically observed gene flux-dependent upper bound on the genome rearrangement rate. In addition, the model developed here is arguably more general than most of the previous ones because it does not

assume the sign or magnitude of the epistasis and the distribution of mutation effects a priori. Negative epistasis and predominantly deleterious mutations arise naturally as a consequence of the additive fitness assumption and a fixed distribution of fitness effects in the reservoir from which new alleles arrive.

Empirical data show that the upper bound of the observed genome rearrangement rates scales sublinearly with the gene flux, whereas many genomes have rearrangement rates much lower than the maximum (fig. 1). This observation implies that microbial genomes evolve in a gene flux regime where the penalty for the higher than optimal recombination rate is greater than that for the lower than optimal rate. The fitness versus recombination rate dependencies tend to display such behavior at lower gene fluxes (fig. 2). Thus, the typical gene flux in prokaryotes appears to be low enough for a certain level of recombination to provide substantial evolutionary benefit.

Conclusions

To summarize, we found that the genome rearrangement rate in bacteria and archaea exhibits an upper bound that scales sublinearly with the gene flux. The genome rearrangement level appears to reflect the rate of large-scale segmental recombination that swaps genomic regions among organisms in the same population, whereas gene flux involves gene-scale allele replacement by genes acquired from a much broader reservoir, the supergenome (Puigbo et al. 2014). In an attempt to account for the observed link between these phenomena, we analyzed a stochastic population model with two independent mechanisms for changing the genome content: 1)

Recombination, which copied random genomic regions between organisms in the population, and 2) gene flux, which introduced novel alleles. The model included the additive fitness and infinite allele assumptions, resulting in fluctuating negative epistasis. Under these assumptions, we found that the steady-state fitness was a nonmonotonic function of the recombination rate, with a pronounced peak at intermediate rates. In agreement with the empirical observations, the optimal recombination rate in the model scaled sublinearly with the gene flux. It should be emphasized that the nonmonotonic effect of recombination is a consequence of the negative epistasis which itself depends on the additive fitness assumption. Indeed, in numerical experiments under a multiplicative fitness model, we observed a monotonic effect of recombination. In our model, when the recombination rate was allowed to evolve, its distribution reached a steady state and the variance in steady state decreased with the increasing gene flux, also in agreement with the observations. Taken together, these findings suggest that the observed link between genome rearrangement and gene flux is maintained by selection for the optimal recombination rate.

Supplementary Material

Supplementary file S1 and figures S1–S5 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Barton NH. 1995. A general model for the evolution of recombination. Genet Res. 65:123–145.
Barton NH. 2010. Mutation and the evolution of recombination. Philos Trans R Soc Lond B Biol Sci. 365:1281–1294.
Barton NH, Charlesworth B. 1998. Why sex and recombination? Science 281:1986–1990
Bodmer WF, Parsons PA. 1962. Linkage and recombination in evolution. Adv Genet. 11:1–100.
Charlesworth B. 1976. Recombination modification in a fluctuating environment. Genetics 83:181–195.
Chicurel M. 2001. Genetics Can organisms speed their own evolution? Science 292:1824–1827.
Christiansen FB, Otto SP, Bergman A, Feldman MW. 1998. Waiting with and without recombination: the time to production of a double mutant. Theor Popul Biol. 53:199–215.
Darmon E, Leach DR. 2014. Bacterial genome instability. Microbiol Mol Biol Rev. 78:1–39.
Draghi JA, Parsons TL, Wagner GP, Plotkin JB. 2010. Mutational robustness can facilitate adaptation. Nature 463:353–355.

Earl DJ, Deem MW. 2004. Evolvability is a selectable trait. Proc Natl Acad Sci U S A. 101:11531–11536.
Eisen JA, Heidelberg JF, White O, Salzberg SL. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. 1:RESEARCH0011.
Eshel I, Feldman MW. 1970. On the evolutionary effect of recombination. Theor Popul Biol. 1:88–100.
Feldman MW, Christiansen FB, Brooks LD. 1980. Evolution of recombination in a constant environment. Proc Natl Acad Sci U S A. 77:4838–4841.
Felsenstein J. 1965. The effect of linkage on directional selection. Genetics 52:349–363.
Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78:737–756.
Felsenstein J, Yokoyama S. 1976. The evolutionary advantage of recombination. II. Individual selection for recombination. Genetics 83:845–859.
Fisher RA. 1930. The genetical theory of natural selection. Oxford: Clarendon Press.
Friedberg EC, et al. 2005. DNA repair and mutagenesis. Washington (DC): ASM Press.
Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43:D261–D269.
Gandon S, Otto SP. 2007. The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. Genetics 175:1835–1853.
Guttman DS, Dykhuizen DE. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science 266:1380–1383.
Hartfield M, Otto SP, Keightley PD. 2012. The maintenance of obligate sex in finite, structured populations subject to recurrent beneficial and deleterious mutation. Evolution 66:3658–3669.
Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8:269–294.
Iles MM, Walters K, Cannings C. 2003. Recombination can evolve in large finite populations given selection on sufficient loci. Genetics 165:2249–2258.
John S, Jain K. 2015. Effect of drift, selection and recombination on the equilibrium frequency of deleterious mutations. J Theor Biol. 365:238–246.
Jones AG, Arnold SJ, Burger R. 2007. The mutation matrix and the evolution of evolvability. Evolution 61:727–745.
Keightley PD, Otto SP. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443:89–92.
Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. Science 332:1193–1196.
Kirschner M, Gerhart J. 1998. Evolvability. Proc Natl Acad Sci U S A. 95:8420–8427.
Kondrashov AS. 1993. Classification of hypotheses on the advantage of amphimixis. J Hered. 84:372–387.
Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36:6688–6719.
Kouyos RD, Silander OK, Bonhoeffer S. 2007. Epistasis between deleterious mutations and the evolution of recombination. Trends Ecol Evol. 22:308–315.
Lehman J, Stanley KO. 2013. Evolvability is inevitable: increasing evolvability without the pressure to adapt. PLoS One 8:e62186.
Lobkovsky AE, Wolf YI, Koonin EV. 2014. Estimation of prokaryotic supergenome size and composition from gene frequency distributions. BMC Genomics 15(Suppl 6):S14.
Martin G, Otto SP, Lenormand T. 2006. Selection for recombination in structured populations. Genetics 172:593–609.

Maynard Smith J. 1976. The evolution of sex. Cambridge: Cambridge University Press.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 106:2–9.

Nei M. 1967. Modification of linkage intensity by natural selection. Genetics 57:625–641.

Novichkov PS, Ratnere I, et al. 2009. ATGC: a database of orthologous genes from closely related prokaryotic genomes and a research platform for microevolution of prokaryotes. Nucleic Acids Res. 37:D448–D454.

Novichkov PS, Wolf YI, Dubchak I, Koonin EV. 2009. Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. J Bacteriol. 191:65–73.

Otto SP, Barton NH. 2001. Selection for recombination in small populations. Evolution 55:1921–1931.

Otto SP, Feldman MW. 1997. Deleterious mutations, variable epistatic interactions, and the evolution of recombination. Theor Popul Biol. 51:134–147.

Otto SP, Lenormand T. 2002. Resolving the paradox of sex and recombination. Nat Rev Genet. 3:252–261.

Partridge L, Barton NH. 2000. Evolving evolvability. Nature 407:457–458.

Peng W, Levine H, Hwa T, Kessler DA. 2004. Analytical study of the effect of recombination on evolution via DNA shuffling. Phys Rev E Stat Nonlin Soft Matter Phys. 69:051911.

Peters AD, Lively CM. 1999. The red queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. Am Nat. 154:393–405.

Puigbo P, Lobkovsky AE, Kristensen DM, Wolf YI, Koonin EV. 2014. Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. BMC Biol. 12:66.

Pylkov KV, Zhivotovsky LA, Feldman MW. 1998. Migration versus mutation in the evolution of recombination under multilocus selection. Genet Res. 71:247–256.

Raeside C, et al. 2014. Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. MBio 5:e01377–14.

Rouzine IM, Coffin JM. 2010. Multi-site adaptation in the presence of infrequent recombination. Theor Popul Biol. 77:189–204.

Roze D. 2009. Diploidy, population structure, and the evolution of recombination. Am Nat. 174(Suppl 1):S79–S94.

Sturtevant A, Mather E. 1938. The interrelations of inversions, heterosis and recombination. Am Nat. 72:447–452.

Takeuchi N, Kaneko K, Koonin EV. 2014. Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. G3 (Bethesda) 4:325–339.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. Science 278:631–637.

Toussaint M. 2003. On the evolution of phenotypic exploration distributions. In: Cotta C, De Jong K, Poli R, Rowe J, editors. Foundations of genetic algorithms 7. New York: Morgan Kaufmann.

Weinreich DM, Watson RA, Chao L. 2005. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution 59:1165–1174.

Weissman DB, Barton NH. 2012. Limits to the rate of adaptive substitution in sexual populations. PLoS Genet. 8:e1002740.

Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV. 2001. Genome alignment, evolution of prokaryotic genome organization and prediction of gene function using genomic context. Genome Res. 11:356–372.

Associate editor: Bill Martin