IMS Collections
Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz
Vol. 4 (2008) 315–324

DOI: 10.1214/074921708000000453

A Duality Identity between a Model of Bacterial Recombination and the Wright–Fisher Diffusion

Xavier Didelot,[1] Jesse E. Taylor,[2] and Joseph C. Watkins[3,*]

University of Oxford and University of Arizona

Abstract: In this article, we establish, using a duality argument, an identity stating that the Laplace transform of the length of a contiguous bacterial recombination region equals the probability of choosing a given allele in a stationary population evolving according to the one-dimensional Wright–Fisher diffusion model. Beyond giving us an improved inferential strategy for parameter estimation in bacterial recombination, the matching of the selection and recombination parameters in the identity also suggests the existence of an intriguing formal relationship between gene conversion and the ancestral selection graph.

1. Introduction

Bacterial genomes are made up of one or a handful of chromosomes which are usually circular, with size ranging from 140 Kb for the endosymbiont *Carsonella* (Nakabachi et al. 2006) to over 12 Mb for the myxobacterium *Sorangium cellulosum* (Pradella et al. 2002). Recombination is not obligate in bacteria but has been shown to happen frequently in nature in a variety of species (e.g. Maynard Smith et al. 1993, Guttman and Dykhuizen 1994, Feil et al. 2001). Bacterial sex is analogous to gene-conversion rather than crossing-over, in the sense that there are always a clear recipient and a clear donor cell, and the resulting bacterium has the same DNA as the recipient for all of its genome except for a small contiguous segment where it is identical to the donor. The average tract length of the imported regions has previously been estimated to be of the order of 1000 bp in several species (Milkman and Bridges 1990, Jolley et al. 2005, Fearnhead et al. 2005). When modeling recombination, the tract length distribution is usually assumed to be geometric (or exponential) with a mean estimated from the data (Falush et al. 2001, McVean et al. 2002, Falush et al. 2003, Suchard et al. 2003). The same assumption is usually made when modeling gene-conversion in eukaryotes (Wiuf and Hein 2000, Frisse et al. 2001).

*Completed during JCW's visit to University of Oxford. JCW would like to take this opportunity to acknowledge the Department of Statistics and the Mathematical Genetics group for their hospitality. This research was supported in part by National Science Foundation grant BCS-0432262.

[1]Department of Statistics, 1 South Parks Road, University of Oxford, Oxford OX1 3TG, UK, e-mail: didelot@stats.ox.ac.uk

[2]Department of Statistics, 1 South Parks Road, University of Oxford, Oxford OX1 3TG, UK, e-mail: jtaylor@stats.ox.ac.uk

[3]Joseph C. Watkins, Department of Mathematics, University of Arizona, 617 North Santa Rita Road, Tucson, Arizona 85716, USA, e-mail: jwatkins@math.arizona.edu

AMS 2000 subject classifications: Primary 92D10; secondary 60K25

Keywords and phrases: bacterial recombination, gene conversion, ancestral selection graph, Wright–Fisher diffusion, M/M/∞ queue, duality identity

Most previous methods estimating the recombination rate and tract length assume that each imported region on the genome is due to exactly one recombination event. However, as the rate of recombination, the average tract length and the time of exposure to recombination increase, so does the probability that several recombination events overlap, meaning that the intersection of chromosomal positions affected by at least two recombination events is non-empty (cf. Figure 1). If the sequences imported by different recombination events can not be distinguished (for example in the case of inter-population recombination as in Falush et al. (2003), it is possible to observe which regions of the genome have been imported (shown in grey on Figure 1), but not the exact starting point and tract length of individual recombination events. Thus, to do inference on the parameters governing the recombination process itself, we need to know how the distribution of length of contiguous imported regions is related to the initiation rate and tract length distribution of individual recombination events. In particular the derivations described here are used in the computer package `ClonalFrame` which infers bacterial microevolution using multilocus sequence data (Didelot and Falush 2007).

Here we consider the genome to be continuous and of size L. Let $\rho/2$, μ^{-1} and δ denote the rate of recombination per genome, average recombination tract length and time of exposure of the genome to recombination respectively. We assume that recombination is uniformly likely to be initiated at any position of the genome, so that the rate of initiation is $\lambda = \frac{\rho}{2}\frac{\delta}{L}$, and that the tract length of a single recombination event is exponentially distributed with mean μ^{-1}. We will also suppose that recombination events are initiated at their upstream boundaries and then proceed downstream; an equivalent result would be obtained if we allowed events to be initiated downstream and proceed upstream or even if we allowed the orientation to be determined at random but independently of all other recombination events. We would obtain a different process if we allowed recombination to proceed in both directions from the initiation point. Because we are concerned with genomes which are large in comparison with the total amount of material likely to have been imported, we will ignore the possibility of wrap-around recombination events in circular genomes or edge effects in linear genomes. With these assumptions in mind, we can model the distribution of imported material in the genome as being generated by a Poisson point process on $\mathbb{R}$, with intensity measure $\lambda\,\mathrm{d}x$, which determines the location of the recombination initiation points, each of which is the left end point of an interval of exponentially distributed length. Our problem is to determine the distribution of the length of maximally overlapping intervals as a function of λ and μ.

To do so, we first observe that this interval-valued process is related to a queue. Indeed, each recombination event (interval) can be identified with a customer who stays in the queue for an exponentially distributed period of time with mean μ^{-1}. Since prior recombination events do not alter the tract length of subsequent recombination events starting in the same region, the queue can be said to have an infinite number of servers. Using this analogy, the distribution of the length of imported regions is the same as that of the busy period of an $\mathrm{M/M/\infty}$ queue (i.e. the contiguous periods of time when there is at least one customer in the system) with arrival rate λ and mean service time requirement μ^{-1}. The length of non-imported regions is distributed as the idle periods of that same queue (i.e. the contiguous periods of time when there is no customer in the system).

Figure 2 shows the cumulative density functions of the busy periods of an $\mathrm{M/M/\infty}$ queue for different values of λ/μ estimated using Monte-Carlo simulations. Although the Laplace transform of the busy period queue has been determined and

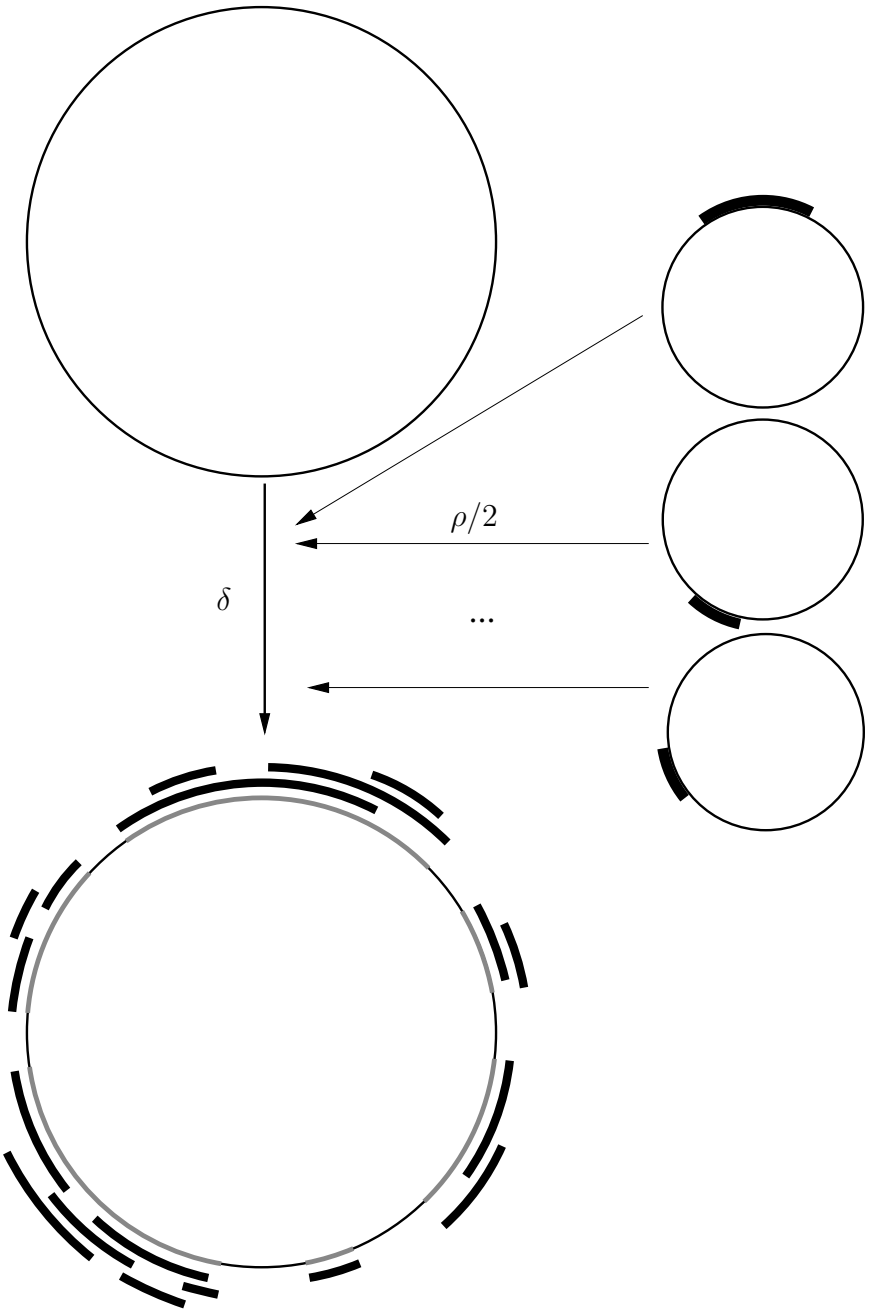

FIG 1. *Illustration of the effect of bacterial recombination. The circle represents the bacterial genome and the bold arcs around it represent the different recombination events. The fragments of the genome affected by recombination are in grey.*

developed by Guillemin and Simonian (1995) and Preater (1997) and is the subject of recent work by Roijers et al. (2006), we provide an alternative derivation of their results which exploits the method of duality. This approach is of interest because it proceeds via the Wright–Fisher diffusion and its moment dual, two stochastic processes which are at the heart of theoretical population genetics.

2. M/M/∞ queues and the Wright–Fisher diffusion

We first recall that the M/M/∞ queue with arrival rate λ and mean waiting time μ^{-1} is a Markov process $(M_t, t \geq 0)$ with state space $\mathbb{N} = \{0, 1, 2, \cdots\}$ and generator

$$G_M\phi(n) = n\mu\big(\phi(n-1) - \phi(n)\big) + \lambda\big(\phi(n+1) - \phi(n)\big), \tag{1}$$

for any bounded function $\phi : \mathbb{N} \to \mathbb{R}$. Because M_t satisfies the strong Markov property, the busy period has the same distribution as the stopping time $\tau_M = \inf\{t > 0 : M_t = 0\}$ if $M_0 = 1$. To determine the Laplace transform of τ_M, we will exploit the fact that a simple time change of M_t leads to a moment dual for the Wright–Fisher diffusion.

To see that this is true, let $(p_t, t \geq 0)$ be a Wright–Fisher diffusion with state space $[0, 1]$ and generator

$$G_p\phi(p) = \frac{1}{2}p(1-p)\phi''(p) + \big(\nu_1(1-p) - \nu_2 p - \sigma p(1-p)\big)\phi'(p). \tag{2}$$

for any twice continuously differentiable function $\phi : [0, 1] \to \mathbb{R}$. As shown in Ethier and Kurtz (1986, Chapter 10), this diffusion process arises as the weak limit of a sequence of suitably scaled Markov chains which model the effects of genetic drift, mutation, and selection on the relative frequency $p \in [0, 1]$ of an allele A_1 in a finite population segregating two alleles, A_1 and A_2. On the diffusive time scale we assume that A_1 mutates to A_2 at rate ν_2, that A_2 mutates to A_1 at rate ν_1, and that A_2 has selective advantage $\sigma \geq 0$ over A_1. We also note that if $\nu_1 > 0$ and $\nu_2 > 0$, then p_t has a unique stationary distribution with density

$$\pi(p) = Cp^{2\nu_1 - 1}(1-p)^{2\nu_2 - 1}e^{-2\sigma p}, \tag{3}$$

where C is a normalizing constant (Ethier and Kurtz, Chapter 10, Lemma 2.1).

If we let $(N_t, t \geq 0)$ be a pure-jump Markov process on $\mathbb{N}$ corresponding to the generator

$$\begin{aligned} G_N\phi(n) = \frac{1}{2}n(n-1)(\phi(n-1) - \phi(n)) \\ + n\nu_1(\phi(n-1) - \phi(n)) + n\sigma(\phi(n+1) - \phi(n)), \end{aligned} \tag{4}$$

and we set $f(p, n) = p^n$, then

$$G_p f(p, n) = G_N f(p, n) - \nu_2 n f(p, n). \tag{5}$$

Since all of the terms appearing in Eq. (5) are bounded, Theorem 4.11 and Corollary 4.13 of Ethier and Kurtz (1986, Chapter 4) imply that p_t and N_t are related by the following duality identity:

$$\mathbb{E}_p[p_t^{n_0}] = \mathbb{E}_{n_0}\left[p^{N_t}e^{-\nu_2\int_0^t N_s\,\mathrm{d}s}\right], \tag{6}$$

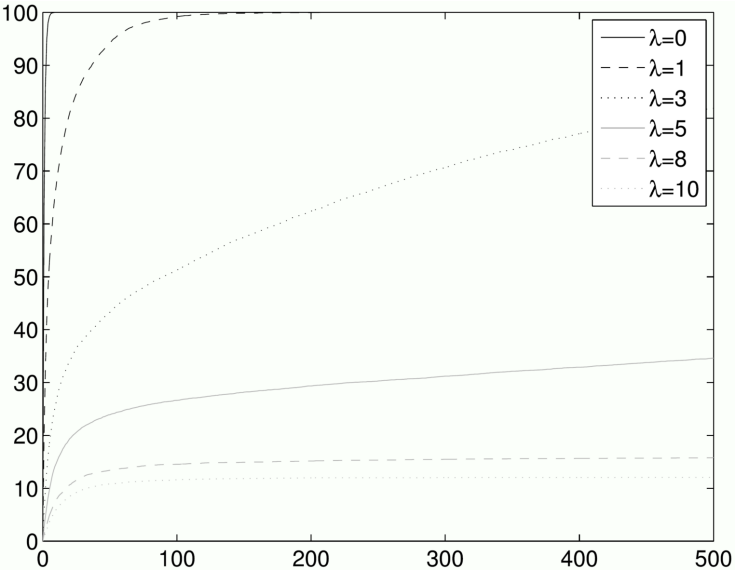

FIG 2. *Cumulative density function of the busy time of an $M/M/\infty$ queue with mean customer requirement $\mu^{-1} = 1$ and different values of the arrival rate λ.*

which holds for all $p \in [0,1]$, all $n_0 \in \mathbb{N}$, and all $t \geq 0$. Furthermore, because N_t almost surely absorbs at 0 at some finite time $T = \inf\{t > 0 : N_t = 0\}$ (see for example Donnelly and Kurtz 1999), it follows that both the right-hand side, and therefore the left-hand side, of Eq. (6) converge as $t \to \infty$. Since p_t takes values in a compact space, this fact, along with the uniqueness of the stationary measure $\pi(p)\,\mathrm{d}p$, implies that the law of p_t tends weakly to $\pi(p)\,\mathrm{d}p$ as $t \to \infty$ and leads to the following equation for the moments of the stationary measure:

$$\int_0^1 p^{n_0}\pi(p)\mathrm{d}p = \mathbb{E}_{n_0}\left[e^{-\nu_2\int_0^T N_s\,\mathrm{d}s}\right]. \tag{7}$$

Now let $T_1, \cdots, T_J$ be the jump times of the process N_t, where $T_J = T$ and $T_0 = 0$, so that J is the number of jumps taken by the process until it absorbs at 0, and let $n_k = N_{T_k}$ be the state occupied by N_t immediately following the k-th jump. Conditional on $n_k = n$, the holding time $T_k - T_{k-1}$ is exponentially distributed with parameter $\frac{1}{2}n(n-1) + n\nu_1 + n\sigma$ and is independent of all $n_j, j \neq k$, and of all other holding times $T_j - T_{j-1}, j \neq k$. Equation (6) can be rewritten as:

$$\begin{aligned}\int_0^1 p^{n_0}\pi(p)\mathrm{d}p &= \mathbb{E}_{n_0}\left[\prod_{k=1}^{J} e^{-\nu_2 n_{k-1}(T_k - T_{k-1})}\right] \\ &\equiv \mathbb{E}_{n_0}\left[\prod_{k=1}^{J} e^{-\nu_2(\tau_k - \tau_{k-1})}\right] \equiv \mathbb{E}_{n_0}\left[e^{-\nu_2\tau}\right],\end{aligned} \tag{8}$$

where, conditional on $n_{k-1} = n$, $\tau_k - \tau_{k-1} \overset{D}{=} n_{k-1}(T_k - T_{k-1})$ is exponentially distributed with parameter $\frac{1}{2}(n-1) + \nu_1 + \sigma$, with the same independence structure as above, and $\tau \equiv \tau_J$.

Introducing the process V_t with generator

$$G_V\phi(n) = \frac{1}{2}(n-1)(\phi(n-1)-\phi(n)) + \nu_1(\phi(n-1)-\phi(n)) + \sigma(\phi(n+1)-\phi(n)), \tag{9}$$

we see that V_t and N_t differ only by a time change, that τ_k is equal in distribution to the time of the k-th jump by V_t, and that $\tau \stackrel{D}{=} \inf\{t > 0 : V_t = 0\}$. It follows from Eq. (8) that

$$\int_0^1 p^{n_0}\pi(p)\mathrm{d}p = \mathbb{E}_{n_0}\left[e^{-\nu_2\tau}\right]. \tag{10}$$

Finally, to relate this result to the original M/M/∞ queue corresponding to (1), observe that by taking $\nu_1 = 1/2$ and $\sigma = \frac{\lambda}{2\mu}$, we will have $M(t) \stackrel{D}{=} V(2\mu t)$ and therefore $\tau_M \stackrel{D}{=} \tau/2\mu$. Thus, if $n_0 = 1$ (so that the beginning of the busy period is initiated by the arrival of a single individual), then from Eqs. (3) and (10) we obtain the following equation for the Laplace transform, $\psi(\alpha)$, of the busy period τ_M:

$$\psi(\alpha) \equiv \mathbb{E}_1\left[e^{-\alpha\tau_M}\right] = \mathbb{E}_1\left[e^{-(\alpha/2\mu)\tau}\right] = \frac{\int_0^1 p(1-p)^{\alpha/\mu-1}e^{-\lambda p/\mu}\,\mathrm{d}p}{\int_0^1 (1-p)^{\alpha/\mu-1}e^{-\lambda p/\mu}\,\mathrm{d}p}. \tag{11}$$

This Laplace transform uniquely specifies the statistical distribution of τ_M. In particular, the moments of τ_M can be simply calculated by evaluating the moment-generating function derivatives $\psi^{(k)}(0) = (-1)^k\mathbb{E}_1\tau_M^k$.

3. Discussion

We now return to the statistical problem of the estimation of the parameters λ and μ^{-1} of the recombination process when all that is known is which regions of the genome have been imported. Since the mean length of non-imported regions is λ^{-1}, the maximum likelihood estimator of λ is the inverse of the mean length of non-imported regions of the bacterial genome. A second parameter estimate can be obtained by taking the inverse Laplace transform of the busy period distribution τ_M and maximizing the likelihood. More generally, we can take the lengths of busy and idle periods and maximize the product of their likelihoods for different value of λ and μ^{-1}.

A simpler approach is to differentiate the Laplace transform and perform a method of moments estimate following the strategy in Section 5 of Roijers et al. (2006). Defining

$$\begin{aligned} I(a,b) &\equiv \int_0^1 (1-p)^{a-1}e^{-bp}\,\mathrm{d}p = e^{-b}\int_0^1 p^{a-1}e^{bp}\,\mathrm{d}p \\ &= e^{-b}\sum_{k=0}^{\infty}\frac{b^k}{k!}\int_0^1 p^{k+a-1}\,\mathrm{d}p = e^{-b}\sum_{k=0}^{\infty}\frac{b^k}{k!}\frac{1}{k+a} = e^{-b}S(a,b), \end{aligned} \tag{12}$$

and noting that

$$\int_0^1 p(1-p)^{a-1}e^{-bp}\,\mathrm{d}p = e^{-b}\int_0^1 (1-p)p^{a-1}e^{bp}\,\mathrm{d}p = I(a,b) - I(a+1,b), \tag{13}$$

it follows that

$$\psi(\alpha) = \frac{I(\alpha/\mu, \lambda/\mu) - I(\alpha/\mu + 1, \lambda/\mu)}{I(\alpha/\mu, \lambda/\mu)} = 1 - \frac{S(\alpha/\mu + 1, \lambda/\mu)}{S(\alpha/\mu, \lambda/\mu)}. \tag{14}$$

To find the expected duration of the busy period, we use Equations 12 and 14 to calculate

$$\psi'(\alpha) = -\frac{\partial_\alpha S(\alpha/\mu + 1, \lambda/\mu)}{S(\alpha/\mu, \lambda/\mu)} + \frac{\partial_\alpha S(\alpha/\mu, \lambda/\mu)}{(S(\alpha/\mu, \lambda/\mu))^2} S(\alpha/\mu + 1, \lambda/\mu). \tag{15}$$

For the first term, the numerator is bounded and denominator is $O(1/\alpha)$ as $\alpha \to 0$. Thus, this term has limit 0. For the fraction in the second term, the singular part in the numerator for α small is $-\mu/\alpha^2$ and in the denominator, it is $(\mu/\alpha)^2$. Thus, this fraction has limit $-1/\mu$ as $\alpha \to 0$. Consequently,

$$\mathbb{E}_1 \tau_M = -\psi'(0) = \frac{1}{\mu} S(1, \lambda/\mu) = \frac{1}{\mu} \sum_{k=0}^{\infty} \frac{(\lambda/\mu)^k}{(k+1)k!} = \frac{e^{\lambda/\mu} - 1}{\lambda} \tag{16}$$

This identity allows us to use the sample mean of the busy period to estimate the average recombination tract length μ^{-1} (Didelot and Falush 2007). Alternatively, this expression can be derived from the detailed balance condition satisfied by the stationary distribution of the queue.

The Wright–Fisher diffusion describes the forward evolution of a population subject to random genetic drift, mutation and selection. The moment duality established in the previous section is closely related to the ancestral selection graph (ASG, Krone and Neuhauser 1997, Neuhauser and Krone 1997), which characterizes the genealogy of a sample of genes collected from such a population. In particular, the duality calculation matches the selective advantage to a term proportional to the recombination rate.

An example of an ancestral selection graph is shown in Figure 3: looking back in time, when k lineages are present, the rate of coalescence is $k(k-1)/2$ as in the coalescent (Kingman 1982) and the rate of branching is $\sigma k/2$. Coalescence represents two lineages finding a common ancestor (represented by the two lines merging on the graph) and branching accounts for the unobserved selective deaths (represented by a lineage splitting into two on the graph). Exactly the same rates of coalescence and branching occur when considering recombination instead of selection and the resulting graph is then called an ancestral recombination graph (ARG, Hudson 1983, Griffiths and Marjoram 1996). In the ARG, the rate of branching per lineage is usually denoted $\rho/2$ and branchings represent recombination events through which a lineage inherits ancestral material from two parents.

Eq. (7) has the following genealogical interpretation. Observe that $\int_0^1 p^{n_0} \pi(p) \mathrm{d}p$ is the probability that a sample of size n_0 drawn from a stationary population evolving according to the Wright–Fisher diffusion consists only of individuals with allele A_1. The process N_t can be thought of as a lines-of-descent modification of the ancestral selection graph (ASG, Krone and Neuhauser 1997), in which n lineages, all of type A_1, are subject to the following events: pairs of lines coalesce at rate $1/2$, individual lines each undergo selective branching at rate σ, and each survive until the most recent (forwards-in-time) A_2-to-A_1 mutation at rate ν_1. The resulting disconnected graph is a subgraph of the ASG and does not contain complete information about the genealogy of the sample. The $e^{-\nu_2 n_k (T_{k-1} - T_k)}$ terms in Eq.

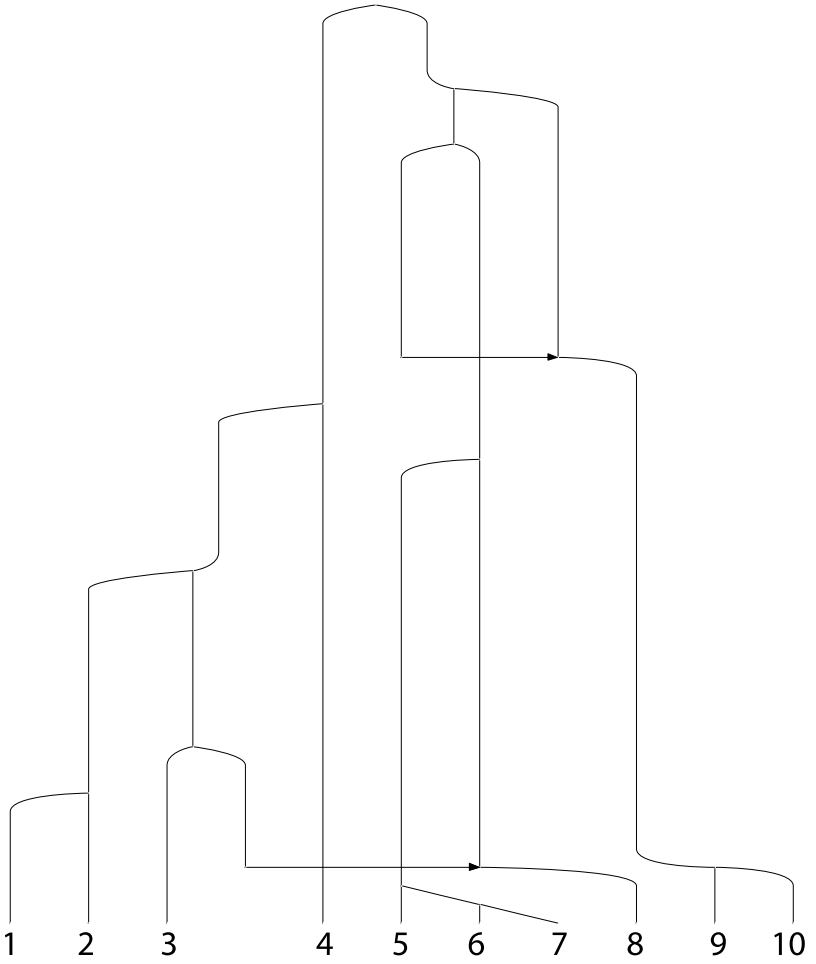

FIG 3. *An example of ancestral recombination/selection graph for* $n = 10$ *individuals and a rate of recombination/selection of* $\rho/2 = \sigma/2 = 1$. *Horizontal arrows represent donor/incoming branches.*

(8) arise from our assumption that the sample consists of A_1 individuals only and that the lines-of-descent survive only until the most recent mutation from A_2; there is a factor of n_k in the exponent because there are n_k lines-of-descent between time T_{k-1} and T_k. Finally, τ is just the extinction time for the lines-of-descent process, i.e., the time when all lineages, ancestral and virtual, have been absorbed by A_2 individuals.

While it is intriguing that in attempting to understand a phylogenetic model of evolution describing the effects of gene conversion on genomic diversification we have been led to a population genetical model of evolution of a non-neutral allele, further research will be necessary to determine whether this is coincidental or hints at some deeper connection. One good starting point would be the particle models of Donnelly and Kurtz (1999) that incorporate both ancestral selection graphs and ancestral recombination graphs into an ancestral inference graph.

References

[1] Didelot, Xavier and Falush, Daniel (2007). Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175** 1251–1266.

[2] Donnelly, P. and Kurtz, T. G. (1999). Genealogical processes for Fleming–Viot models with selection and recombination. *Ann. Appl. Prob.* **9** 1091–1148.

[3] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. John Wiley & Sons, New York.

[4] Falush, D., Kraft, C., Taylor, N. S., Correa, P., Fox, J. G., Achtman, M. and Suerbaum, S. (2001). Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size, and minimal age. *Proc. Natl. Acad. Sci. USA* **98** 15056–15061.

[5] Falush, D., Stephens, M. and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164** 1567–1587.

[6] Fearnhead, P., Smith, N. G., Barrigas, M., Fox, A. and French, N. (2005). Analysis of recombination in *Campylobacter jejuni* from MLST population data. *J. Mol. Evol.* **61** 333–340.

[7] Feil, E. J., Holmes, E. C., Enright, M. C., Bessen, D. E., Day, N. P. J., Chan, M.-S., Hood, D. W., Zhou, J. and Spratt, B. G. (2001). Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. USA* **98** 182–187.

[8] Frisse, L., Hudson, R. R., Bartoszewicz, A., Wall, J. D., Donfack, J. and Di Rienzo, A. (2001). Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69** 831–843.

[9] Griffiths, R. C. and Marjoram. P. (1996). Ancestral inference from samples of DNA sequences with recombination. *J. Computational Biology* **3** 479–502.

[10] Guillemin, F. and Simonian, A. (1995). Transient characteristics of an $M/M/\infty$ system. *Advances in Applied Probability* **27** 862–888.

[11] Guttman, D. S. and Dykhuizen, D. E. (1994). Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266** 1380–1383.

[12] Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology* **23** 183–201.

[13] Jolley, K. A., Wilson, D. J., Kriz, P., McVean, G. and Maiden, M. C. J. (2005). The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*. *Mol. Biol. Evol.* **22** 562–569.

[14] Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications* **13** 235–248.

[15] Krone, S. M. and Neuhauser, C. (1997). Ancestral process with selection. *Theor. Pop. Biol.* **51** 210–237.

[16] Maynard Smith, J., Smith, N. H., O'Rourke, M. and Spratt, B. (1993). How clonal are bacteria? *PNAS* **90** (10) 4384–4388.

[17] McVean, G., Awadalla, P. and Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **2002** 1231–1241.

[18] Milkman, R. and Bridges, M. M. (1990). Molecular Evolution of the Escherichia coli Chromosome. III. Clonal Frames. *Genetics* **126** 505–517.

[19] Nakabachi, Atsushi, Yamashita, Atsushi, Toh, Hidehiro, Ishikawa, Hajime, Dunbar, Helen E., Moran, Nancy A. and Hattori, Masahira (2006). The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314** 267.

[20] Neuhauser, C. and Krone, S. M. (1997). The genealogy of samples in models with selection. *Genetics* **145** 519–534.

[21] Pradella, S., Hans, A., Spröer, C., Reichenbach, H., Gerth, K. and Beyer, S. (2002). Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. *Arch. Microbiol.* **178** 484–492.

[22] Preater, J. (1997). M/M/∞ transience revisited. *Journal of Applied Probability* **34** 1061–1067.

[23] Roijers, F., Mandjes, M. and van den Berg, H. (2007). Analysis of congestion periods of an M/M/∞-queue. *Performance Evaluation* **64** 737–754.

[24] Suchard, M. A., Weiss, R. E., Dorman, K. S. and Sinsheimer, J. S. (2003). Inferring spatial phylogenetic variation along nucleotide sequences: A multiple change-point model. *Journal of the American Statistical Association* **98** 427–437.

[25] Wiuf, C. and Hein, J. (2000). The coalescent with gene conversion. *Genetics* **155** 451–462.