

## A Additional analytical results

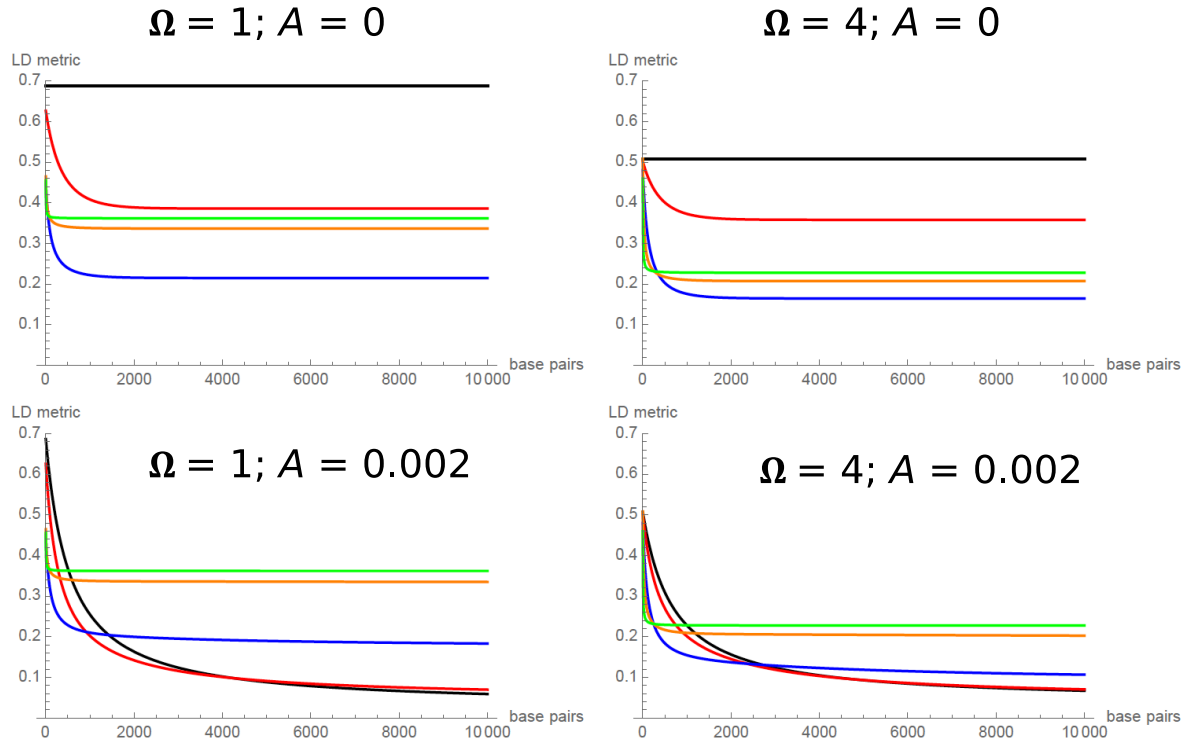


Figure A: As Figure 3(c, d) in the main text but with different rates of sex (left column  $\Omega = 1$ ; right column  $\Omega = 4$ ) .

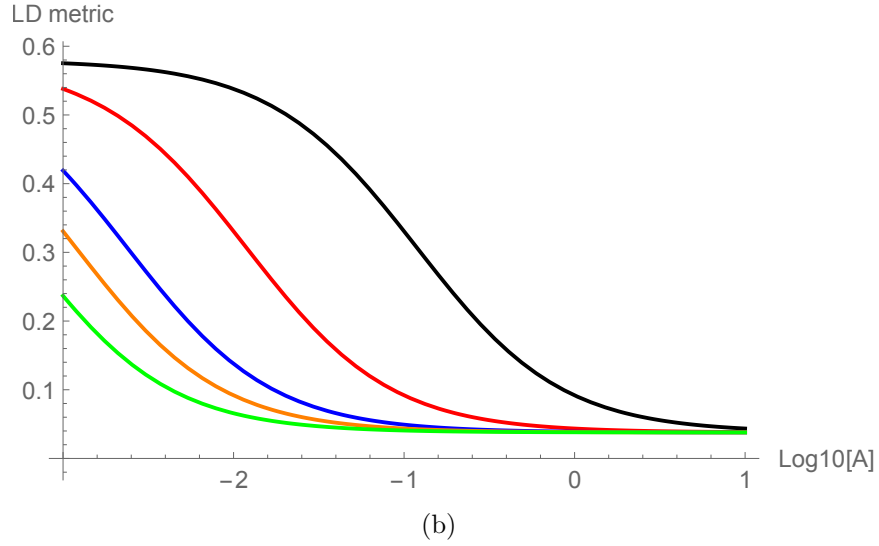
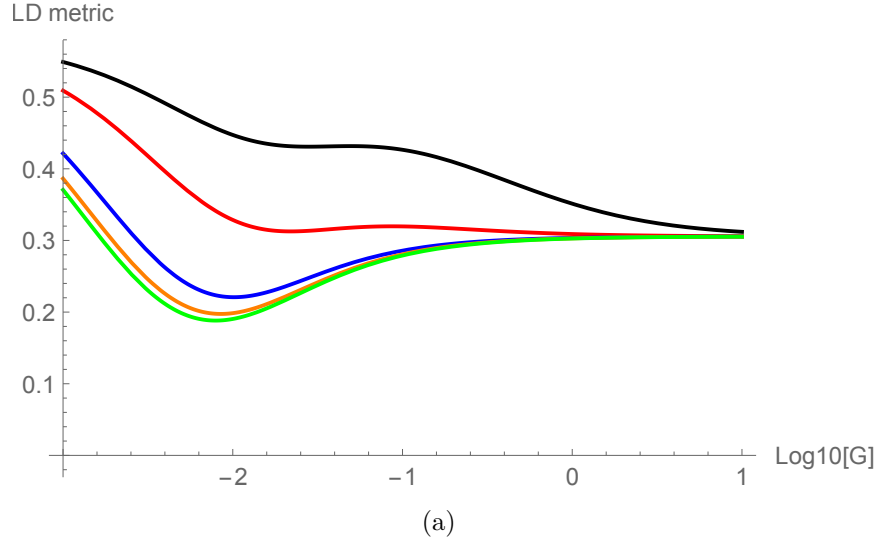


Figure B: The linkage disequilibrium metric  $r_d^2$  as a function of either (a) the scaled rate of mitotic gene conversion initiation  $G$  or (b) the scaled rate of mitotic crossing over  $A$ . Each line corresponds to a different distance  $d$  between sites (black = 10, red = 100, blue = 500, orange = 1000, green = 2000). For (a), the mean gene conversion tract length is  $\lambda = 500$ . Other parameters:  $\Omega = 2$ ;  $\rho = \Gamma_{1S} = 0$ ;  $\rho_A = 0$  in (a) and  $G = 0$  in (b).

## B Testing the simulation program

### (a) Number of non-recombining segments

To ensure that the crossover rate is the same between *FacSexCoalescent* and **ms**, we calculated the mean number of non-recombining segments, minus one (since one segment will always be produced with no crossing over), while assuming obligate sex in *FacSexCoalescent*. 10 samples were simulated assuming a population size  $N = 10,000$ , with  $L = 1,001$  sites per sample. Tables A and B outline the mean number of trees over 1,000 simulations, for different crossover rates and population structure. Results match up between *FacSexCoalescent* and **ms**.

### (b) Expected diversity within and between individuals

We also investigated if the average pairwise diversity over the genetic segment matched theoretical results (Bengtsson 2003; Ceplitis 2003; Hartfield *et al.* 2016). For each simulation run for a particular scenario, polymorphisms are put into one of ten equally-sized bins depending on its position on the chromosome. The within- and between-individual diversity is calculated based on all polymorphisms present in a bin, and the mean taken over all comparisons. This is repeated over all simulation runs. Figures C, D plot the mean of means over 1,000 simulations, with confidence intervals. These results are compared to theoretical expectations  $\theta E[\tau]$ , where  $\theta$  is the population level neutral mutation rate and  $E[\tau]$  the expected pairwise coalescent time in units of  $2N$  generations. Without gene conversion, the expected coalescent times between individuals ( $\tau_b$ ) or within individuals ( $\tau_w$ ) equal:

Table A: Number of minimum recombinants test. Values in brackets represent 95% confidence intervals.

Single Deme, meiotic crossovers only		
Crossover rate, $4Nc$	Mean min. rec. events, FSC	Mean min. rec. events, MS
1	2.85 (0.124)	2.82 (0.125)
2	5.47 (0.193)	5.55 (0.192)
4	11.40 (0.318)	11.38 (0.317)
10	27.70 (0.548)	27.71 (0.563)
20	55.25 (0.842)	54.56 (0.843)
50	130.87 (1.39)	131.10 (1.40)
100	241.47 (1.70)	239.72 (1.79)
Two Demes ( $2N_Tm = 1$ ), meiotic crossovers only		
1	3.53 (0.143)	3.50 (0.138)
2	6.84 (0.224)	6.74 (0.230)
4	13.64 (0.368)	13.89 (0.375)
10	34.16 (0.663)	33.34 (0.659)
20	66.13 (1.03)	66.48 (1.00)
50	156.94 (1.57)	155.23 (1.66)
100	285.02 (1.95)	285.60 (1.96)
Four Demes ( $2N_Tm = 1$ ), meiotic crossovers only		
1	5.74 (0.208)	5.87 (0.220)
2	11.63 (0.347)	11.36 (0.335)
4	23.05 (0.567)	22.8 (0.567)
10	55.71 (1.07)	55.59 (1.06)
20	108.59 (1.56)	107.80 (1.55)
50	243.69 (2.34)	243.18 (2.35)
100	420.36 (2.79)	419.17 (2.71)

Table B: Number of minimum recombinants test. Values in brackets represent 95% confidence intervals.

Single deme, gene conversion, $\lambda = 10$		
Gene conversion rate $4Ng(L - 1)$	Mean min. rec. events FSC	Mean min. rec. events MS
1	5.536 (0.246)	5.66 (0.257)
2	11.20 (0.404)	11.27 (0.436)
4	22.30 (0.703)	22.24 (0.712)
10	53.30 (1.43)	55.85 (1.43)
20	105.72 (2.58)	103.95 (2.56)
30	155.15 (3.38)	152.90 (3.43)
Two Demes ( $2N_Tm = 1$ ), gene conversion, $\lambda = 10$		
1	6.71 (0.293)	6.77 (0.298)
2	13.43 (0.494)	14.04 (0.474)
4	26.81 (0.870)	27.67 (0.853)
10	66.84 (1.78)	65.97 (1.80)
20	128.03 (3.08)	127.70 (2.95)
30	184.76 (3.94)	182.01 (4.05)
Single deme, gene conversion, $\lambda = 200$		
1	5.44 (0.232)	5.56 (0.227)
2	11.30 (0.380)	11.28 (0.390)
4	21.70 (0.582)	22.19 (0.578)
10	54.43 (1.04)	55.18 (1.08)
20	105.67 (1.46)	106.55 (1.47)
30	154.96 (1.82)	152.64 (1.80)

$$\begin{aligned}
E[\tau_b] &= 1 + \frac{1}{\Omega} \\
E[\tau_w] &= 1 + \frac{2}{\Omega}
\end{aligned}
\tag{1}$$

where  $\Omega = 2N_T\sigma$ . Figure C shows results from a high-sex, single deme population with no gene conversion; simulated pairwise diversity matches up to theoretical expectations given by Equation 1.

When both gene conversion and population structure are present, we consider the mean coalescent times either between demes ( $\tau_d$ ), between individuals within demes ( $\tau_b$ ), or within individuals ( $\tau_w$ ):

$$\begin{aligned}
E[\tau_d] &\approx \left( \frac{1 + \phi}{2 + \phi} \right) + \frac{1}{\Gamma_1(2 + \phi)} + \frac{d - 1}{2M} \\
E[\tau_b] &\approx \left( \frac{1 + \phi}{2 + \phi} \right) + \frac{1}{\Gamma_1(2 + \phi)} \\
E[\tau_w] &\approx \left( \frac{\phi}{2 + \phi} \right) + \frac{2}{\Gamma_1(2 + \phi)}
\end{aligned}
\tag{2}$$

Here,  $\Gamma_1 = 4N_T\gamma_1$  is the population-level haploid gene conversion rate at a single site,  $M = 2N_Tm$  the migration rate, and  $\phi = \Omega/\Gamma_1$  the ratio of sex to gene conversion at a single site. Note that while the equations for a single locus (Hartfield *et al.* 2016, Eq. 11) scaled the gene conversion probability by  $2N$ , here the population-level gene conversion rate  $\Gamma_1$  is scaled by  $4N$ . This is because the multi-locus gene-conversion routine considers haploid probabilities of gene conversion, while the single-locus analytical derivations of Hartfield *et al.* (2016) assume diploid gene conversion probabilities, necessitating an extra factor of two when scaling the gene

conversion probability. Second, if  $\Gamma = 4Ng(L - 1)$  is the scaled gene-conversion rate across the entire genetic sample, then  $\Gamma_1 = \Gamma/Q$  for  $Q = (L - 1)/\lambda$ . Figure D shows simulated pairwise diversity results from a low-sex, four-deme population with low rates of sex and mitotic gene conversion. We observe that calculations of pairwise diversity match up to theoretical expectations given by Equation 2.

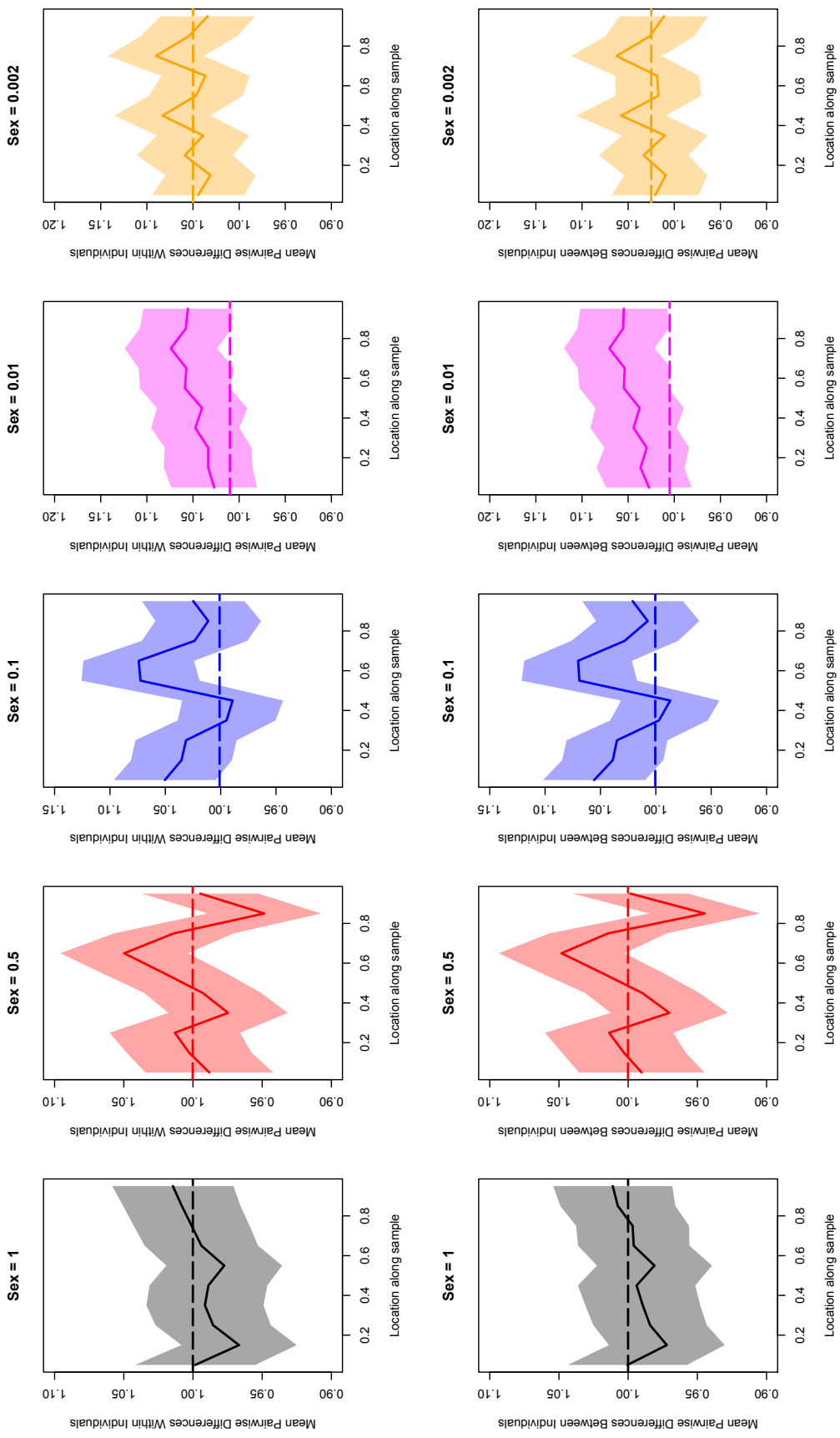


Figure C: Mean within-individual (top row) or between-individual (bottom row) pairwise diversity values in 10 equally-sized bins over the genetic sample, as a function of the location (scaled to 1). Solid line is the mean value over 1,000 simulations; curves represent 95% confidence intervals. Dashed lines are expected results from Equation 2. 50 paired samples were simulated (100 samples in total),  $N = 10,000$ , scaled mutation rate across the whole sample  $\theta = 4N\mu = 10$  (so the per-window  $\theta = 1$  as 10 windows were used), scaled meiotic crossover rate during sex  $4Nc = 40$ . Rates of sex are as denoted in figure headings.



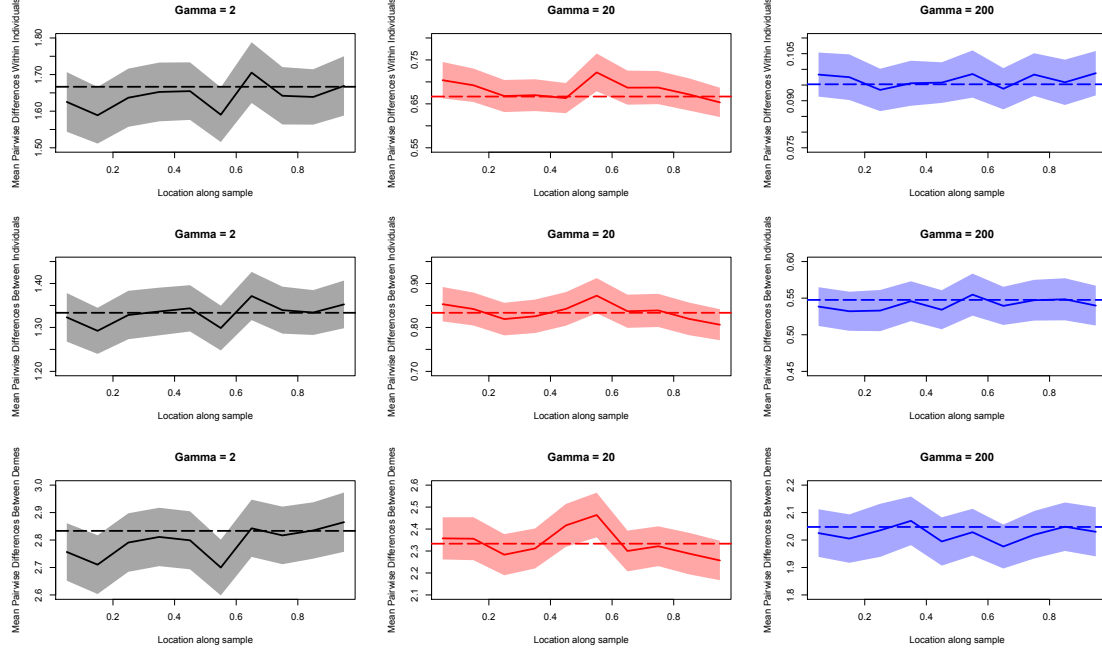


Figure D: Mean within-individual (top row), between-individual, within deme (middle row) or between-deme (bottom row) pairwise diversity values in 10 equally-sized bins over the genetic sample, as a function of the location (scaled to 1). Solid line is the mean value over 1,000 simulations; curves represent 95% confidence intervals. Dashed lines are expected results from Equation 2. 50 paired samples were simulated over 4 demes according to an island model (13 paired samples taken from demes 1 and 2; 12 paired samples taken from demes 3 and 4),  $N_T = 10,000$ , scaled mutation rate  $\theta = 4N\mu = 10$  (with per-window  $\theta = 1$  as 10 windows are used), scaled migration rate  $2N_Tm = 1$ . For fixed rates of sex  $\Omega = 2$ , mitotic gene conversion is set to  $\Gamma = 2$  (black line); 20 (red line); or 200 (blue line) with  $\lambda = 1,000$  over  $L = 10,001$  sites (so the sex-to-gene conversion ratio per site  $\phi = 10, 1$  or  $0.1$ ).

## C Additional Simulation Results

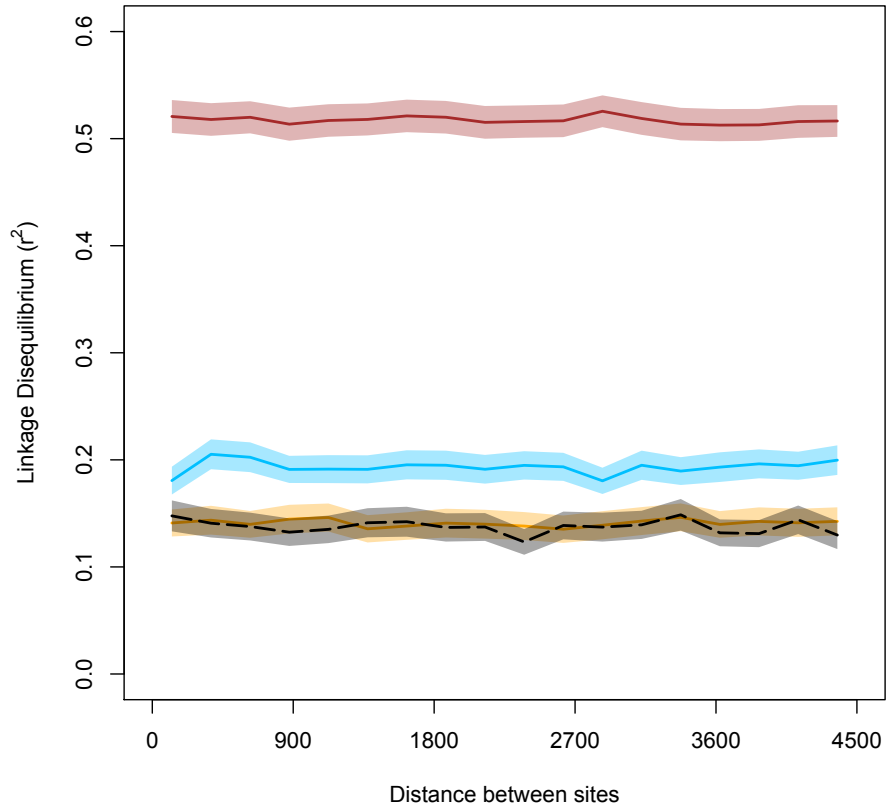
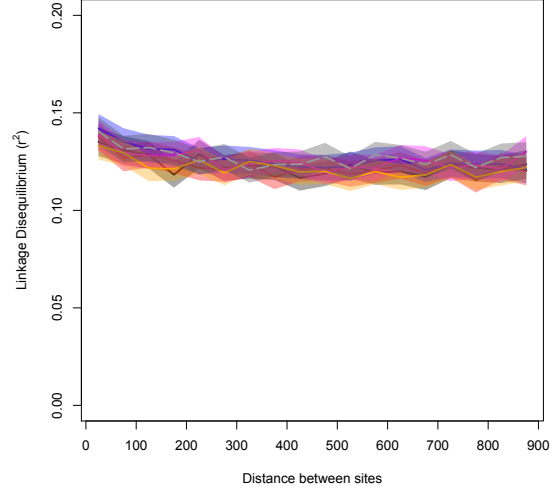
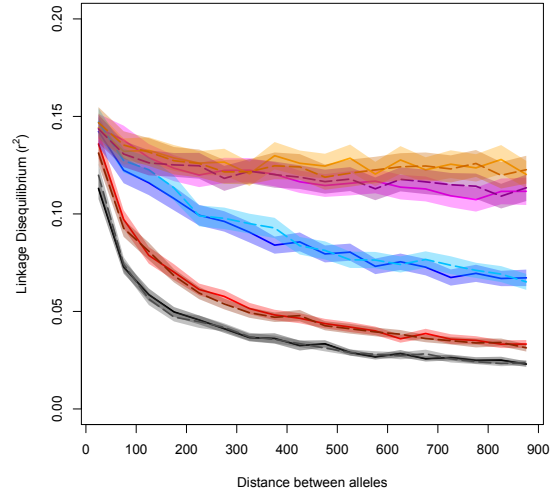


Figure E: Decay of linkage disequilibrium for low rates of sex, where meiotic crossing over is present. Parameters and colours are the same as in Figure 5 in the main text, except results are shown between 125 and 4,375 sites apart.



(a)



(b)

Figure F: Decay of linkage disequilibrium between sites 25 to 875, as a function of distance between two sites. Parameters are the same as Figure 6 in the main text, but with mitotic  $\Gamma = 2$ .

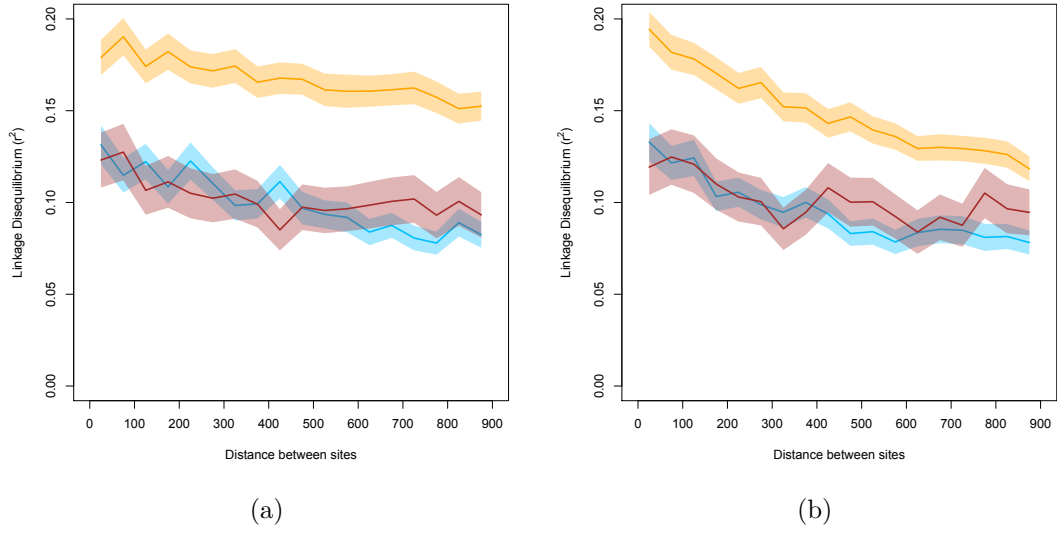


Figure G: Plot of linkage disequilibrium as a function of distance between two sites, where sex is fixed and gene conversion varies. Mitotic recombination is (a) absent of (b) present at rate  $\rho_A = 4N(c_A) = 10$ . Other parameters are the same as Figure 7 in the main text, but  $r^2$  in all plots is shown for distances between 25 to 875 sites apart.

## Gene conversion initiation versus tract length

In the main text, we only examined the effect of different amounts of gene conversion by changing the per-site probability of gene conversion initiation,  $g$  (or its population size scaled equivalents). The total amount of DNA affected by gene conversion depends on the product of initiation probability  $g$  and average tract length  $\lambda$ , so changes in  $g$  and  $\lambda$  are equivalent in this sense. Indeed, the probability of (single-site) coalescence by gene conversion depends only on the product  $g\lambda$ . However, as described in the analytical section,  $g$  and  $\lambda$  differentially affect whether a gene conversion even will break up a haplotype (i.e., they have separate effects on the probabilities  $\gamma_1$  and  $\gamma_2$ ). Thus, we expect changes in  $g$  and  $\lambda$  to differentially affect linkage disequilibrium, even when the total amount of gene conversion ( $g\lambda$ ) is held constant. Hence we also investigated results with varying  $g$  and  $\lambda$  while holding  $g\lambda$  constant. As expected, linkage disequilibrium decays more rapidly for higher  $g$  values with lower  $\lambda$  (Figure H). By increasing  $g$  and proportionally reducing  $\lambda$ , there are more gene conversion events that cover one site but not a neighbouring site, thereby splitting the genealogical histories of the two sites.

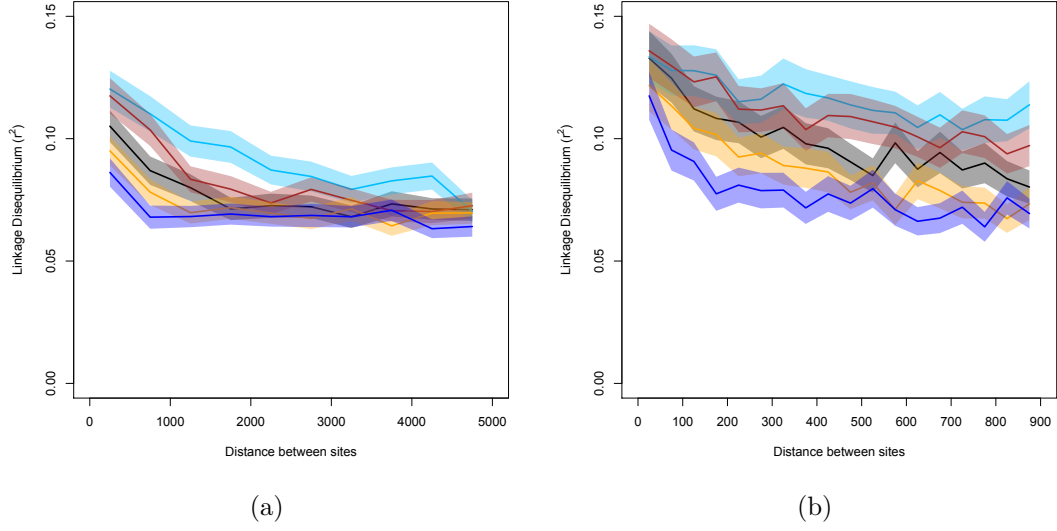


Figure H: (a) Plot of linkage disequilibrium as a function of distance between two sites.  $\Omega = 2$  and  $\Gamma = 5$ ,  $\lambda = 4,000$  (light blue line);  $\Gamma = 10$ ,  $\lambda = 2,000$  (red line);  $\Gamma = 20$ ,  $\lambda = 1,000$  (black line);  $\Gamma = 40$ ,  $\lambda = 500$  (orange line), or  $\Gamma = 80$ ,  $\lambda = 250$  (dark blue line). Colour bands represent 95% confidence intervals. Note that although simulations were performed over a tract of length 10,001 sites, results are only shown for distances between 250 and 4,750 sites apart since curves become indistinguishable at longer distances. (b)  $r^2$  in all plots is shown for distances between 25 to 875 sites apart.

## Effect of Population Subdivision, high frequencies of sex

When sex is not too rare, baseline rates of linkage disequilibrium are higher than those observed in a single-deme population, and the decay in linkage disequilibrium is equivalent to that obtained from an obligate sex population with an effective crossover probability  $c_{eff} = c\sigma$ . Including gene conversion produces qualitatively similar results to the single-population case, albeit with a higher baseline rate of disequilibrium due to population structure (Figure I).

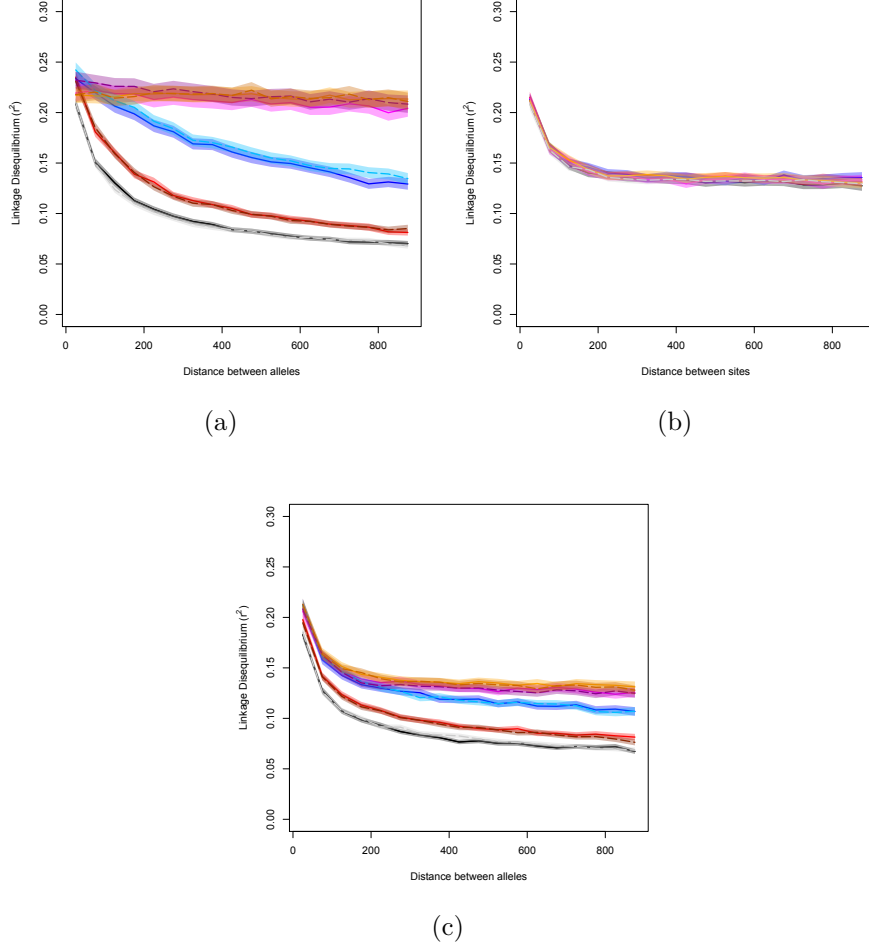


Figure I: (a) Decay of linkage disequilibrium between distances of 25 to 875 sites apart, as a function of distance between two sites. Different colours denote individual rates of sex, as denoted in Figure 4 in the main text. The population is split over 4 demes assuming an island model, with migration rate  $2N_Tm = 1$ . Solid line is the mean value over 1,000 simulations; fainter curves represent 95% confidence intervals. 50 paired samples were simulated (100 samples in total, with 13 paired samples taken from demes 1 and 2 and 12 paired samples taken from demes 3 and 4),  $N_T = 10,000$ , scaled mutation rate  $\theta = 4N_T\mu = 10$ , scaled crossover rate during sex  $4N_Tc = 40$ . Dashed lines are results from obligate sex simulations ran using `ms` using crossover rate  $4N_Tc\sigma$ . (b) No meiotic crossing over but mitotic gene conversion present with rate  $\Gamma = 20$  and average gene conversion tract length  $\lambda = 100$  sites. (c) With both meiotic crossing over (rate during sex  $4N_Tc = 40$ ) and mitotic gene conversion ( $\Gamma = 20$ , average gene conversion tract length  $\lambda = 100$  sites).



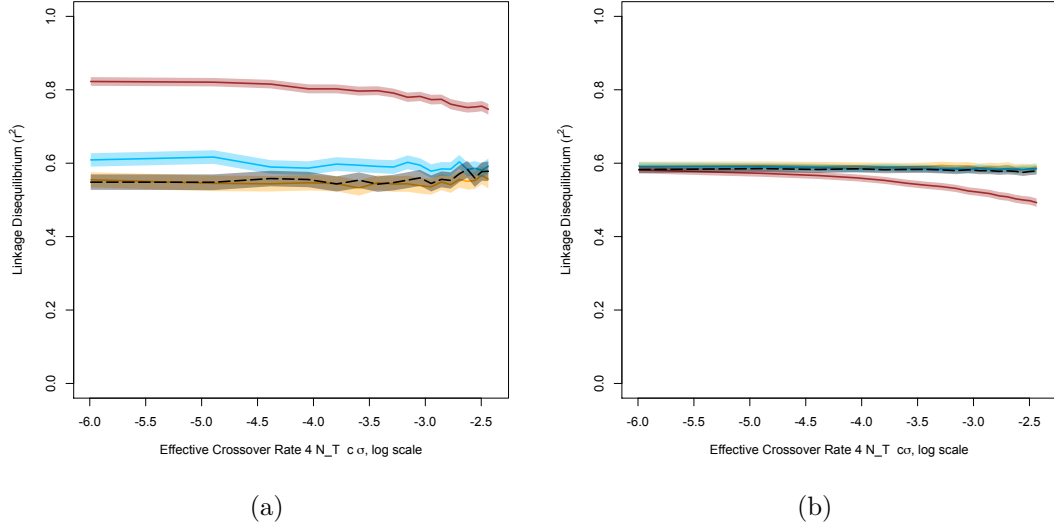


Figure J: Plot of linkage disequilibrium (measured using  $r^2$ ) as a function of the rescaled recombination rate  $4N_T c \sigma$  and after removing polymorphisms with minor allele frequency less than 15%. 50 paired samples were simulated over 4 demes according to an island model (13 paired samples taken from demes 1 and 2; 12 paired samples taken from demes 3 and 4),  $N_T = 10,000$ , scaled mutation rate  $\theta = 4N\mu = 10$ , scaled crossover rate  $4N_T c \sigma = 0.1$ , and scaled migration rate (a)  $2N_T m = 10$  or (b)  $2N_T m = 0.1$ . Each sample contained 100,001 sites, with  $r^2$  plotted over only the first 90,000. Shading around lines indicate 95% confidence intervals.  $r^2$  is shown for distances between 2,500 to 87,500 sites apart.

## Low rate of sex with mitotic gene conversion

We also ran island model simulations where either the rate of sex or gene conversion per site was fixed (at values of  $\Omega = 2$  and  $4N_T g \lambda = 2$  respectively), and set the sex-to-gene-conversion ratio  $\phi$  to either 10, 1 or 0.1. Results are qualitatively similar to the panmictic population case (Figure K). With fixed rates of gene conversion, lower rates of sex increase baseline  $r^2$  values. With fixed rates of sex, high rates of gene conversion also raises  $r^2$ . We can understand this behaviour in terms of how the  $(E[\tau]^2/Var[\tau])$  ratio affects  $r^2$ . Hartfield *et al.* (2016) determined the mean coalescent times for the facultative sex coalescent in an island model, but not the variance. Here, we derive the variance in coalescent times to determine how the ratio  $E[\tau]^2/Var[\tau]$  is affected by population structure. We achieve this by first deriving the associated Laplace transform of the coalescent process, which can be used to derive the variance in coalescent times (Herbots 1997; Hartfield *et al.* 2016). We outline the mathematical argument here, with further details in Section D of Supplementary Mathematica File S1.

Let there be  $d$  demes, each containing  $n$  individuals so the total population size is  $N_T = nd$ . An individual migrates from one deme to another each generation with probability  $m$ ; reproduces sexually with probability  $\sigma$ ; and gene conversion acts at a site in diploid individuals with probability  $\gamma$ . Two samples coalesce within each deme with probability  $1/(2n)$ . If the population size is sufficiently large and each parameter sufficiently small, then we can define the coalescent process acting in continuous time, in time units of  $2N_T$  generations. We define the rescaled parameters  $\Omega = 2N_T \sigma$ ,  $\Gamma = 2N_T \gamma_1$  and  $M = 2N_T m$ . Coalescence then occurs with rate  $d$  (i.e., with mean time  $1/d$ ). Note that the haploid gene conversion

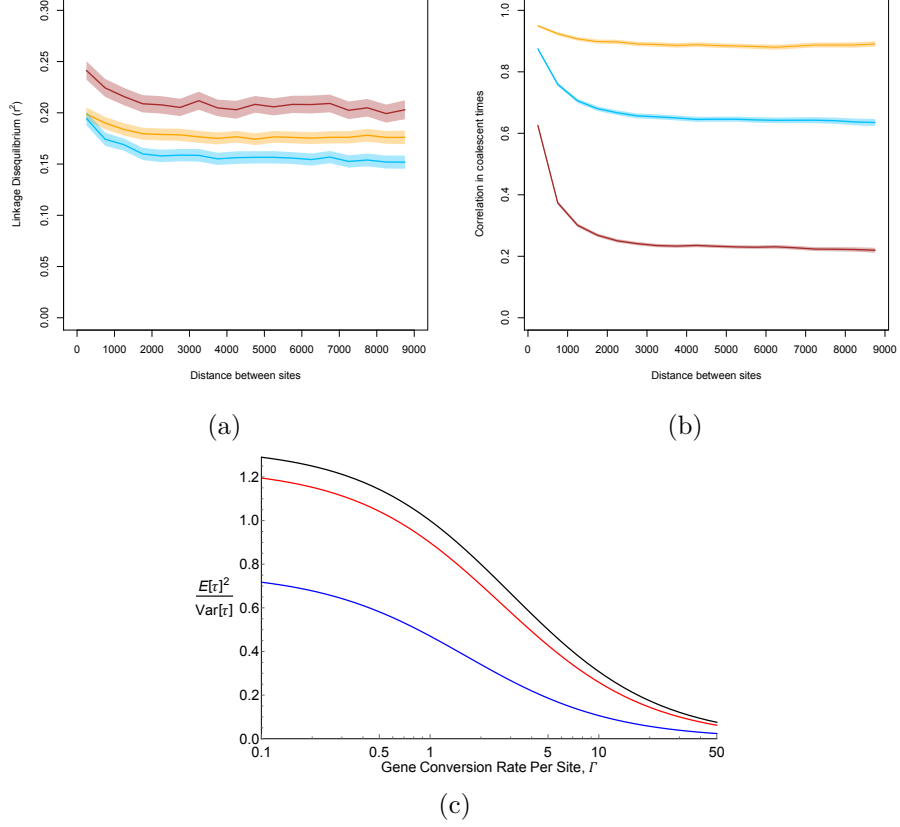


Figure K: (a) Plot of linkage disequilibrium, as measured using  $r^2$ , as a function of distance between two sites. 50 paired samples were simulated over 4 demes according to an island model,  $N_T = 10,000$ , scaled mutation rate  $\theta = 4N\mu = 10$ , scaled migration rate  $2N_Tm = 1$ . Each sample contained 10,001 sites, with  $r^2$  plotted between distances of 250 to 8,750 sites apart. For a fixed rate of sex  $\Omega = 2$ , gene conversion is set to  $\Gamma = 2$  (orange line); 20 (blue line); or 200 (red line) with  $\lambda = 1,000$ . Shading around lines indicate 95% confidence intervals. (b) Correlation in coalescent times between sites. Parameters and line colours are the same as (a). (c)  $(E[\tau]^2/\text{Var}[\tau])$  for within-individual coalescent times, as a function of the per-site gene conversion rate  $\Gamma$ . Results are plotted for the single-deme case with  $\Omega = 2$  (black line), or for the island model with 4 demes, and  $M = 10$  (red line) or 0.1 (blue line).

probabilities can be used below if  $\gamma_1$  is instead scaled by  $4N$ .

Two samples can either be taken from different demes; from different individuals from the same deme; or from different haplotypes with the same individual. Denote the random variables associated with the coalescent times in these states as  $\tau_d$ ,  $\tau_b$ ,  $\tau_w$ . The Laplace transforms for each of these cases are the solutions to the equations:

$$\begin{aligned} E[e^{-s\tau_d}] &= E[e^{-sX_d}]E[e^{-s\tau_b}] \\ E[e^{-s\tau_b}] &= E[e^{-sX_b}] \left( \frac{2M}{2(M+d)} E[e^{-s\tau_d}] + \frac{d}{2(M+d)} (E[e^{-s\tau_w}] + 1) \right) \\ E[e^{-s\tau_w}] &= E[e^{-sX_w}] \left( \frac{\Gamma}{\Gamma + \Omega} + \frac{\Omega}{\Gamma + \Omega} E[e^{-s\tau_w}] \right) \end{aligned} \quad (3)$$

$X_d$ ,  $X_b$  and  $X_w$  are random variables for the waiting times before leaving each state. For the continuous time process, these values are exponentially distributed with rates  $2M/(d-1)$ ;  $2(M+d)$ ; and  $\Omega + \Gamma$  respectively. Hence  $E[e^{-s\tau_*}]$  (where  $*$  =  $d$ ,  $b$  or  $w$ ) equals  $\Lambda/(\Lambda + s)$  for  $\Lambda$  the exponential rate parameter. If two samples from the same deme change state, then one of the two samples migrates into the same deme as the other. Given a state change, this event occurs with probability 1. For two samples in the same deme, if a state change occurs then one sample either migrates to a different deme with probability  $2M/[2(M+d)]$ ; be descended from the same individual with probability  $d/[2(M+d)]$  or coalesce with the same probability. The within-individual state changes are the same as for the model without population structure (Hartfield *et al.* 2016).

We can solve the system of equations to yield solutions to the Laplace transform, although solutions are unwieldy (Supplementary Mathematica File S1). The

variance in coalescent times can be calculated by taking the second derivate of  $E[e^{-s\tau_*}]$  evaluated at  $s = 0$  to obtain  $E[\tau_*^2]$ , then subtracting  $(E[\tau_*])^2$  from these solutions:

$$\begin{aligned} Var[\tau_d] &= \frac{(d-1)^2}{4M^2} + \frac{(d-1)^2(\Gamma + \Omega)}{dM(2\Gamma + \Omega)} + \frac{3 + 2\Omega + (\Gamma + \Omega)^2}{(2\Gamma + \Omega)^2} \\ Var[\tau_b] &= \frac{1 + d(M + d - 2)}{dM} + \frac{(3 - \Gamma)(1 - \Gamma)}{(2\Gamma + \Omega)^2} + \frac{2dM - \Gamma((d-1)^2 + 2dM)}{dM(2\Gamma + \Omega)} \quad (4) \\ Var[\tau_w] &= \frac{1 + d(M + d - 2)}{dM} + \frac{4(1 - \Gamma)}{(2\Gamma + \Omega)^2} + \frac{2dM - 2\Gamma(1 + d(M + d - 2))}{dM(2\Gamma + \Omega)} \end{aligned}$$

Note that if  $d = 1$ ,  $Var[\tau_w]$  and  $Var[\tau_b]$  reduce to the single-population result (Hartfield *et al.* 2016) and  $Var[\tau_d] = Var[\tau_b]$ . The derivative of  $Var[\tau_*]$  with respect to  $M$  is negative; that is, the variance in coalescent times increases if the migration rate is reduced, which will lower the  $(E[\tau]^2/Var[\tau])$  ratio and increase  $r^2$ .

Figure K(a) plots  $r^2$  where the rate of sex is fixed. We also confirmed that the correlation in coalescent times is lower with increased rates of gene conversion, assuming a fixed rate of sex (Figure K(b)). Hence the observed increased in  $r^2$  is due to the lower  $(E[\tau]^2/Var[\tau])$  ratio caused by lower migration rates (Figure K(c)).

## Literature Cited

- Bengtsson, B. O., 2003 Genetic variation in organisms with sexual and asexual reproduction. *J. Evol. Biol.* **16**: 189–199.
- Ceplitis, A., 2003 Coalescence times and the Meselson effect in asexual eukaryotes. *Genet. Res.* **82**: 183–190.
- Hartfield, M., S. I. Wright, and A. F. Agrawal, 2016 Coalescent times and patterns of genetic diversity in species with facultative sex: Effects of gene conversion, population structure, and heterogeneity. *Genetics* **202**: 297–312.
- Herbots, H. M., 1997 The structured coalescent. In *Progress in Population Genetics and Human Evolution*, edited by P. Donnelly and S. Tavaré, volume 87 of *IMA Volumes in Mathematics and its Applications*, pp. 231–255, New York, Springer.