

# **Ancestral Processes with Bias in Ectopic Gene Conversion or Migration**

**Shuhei MANO**

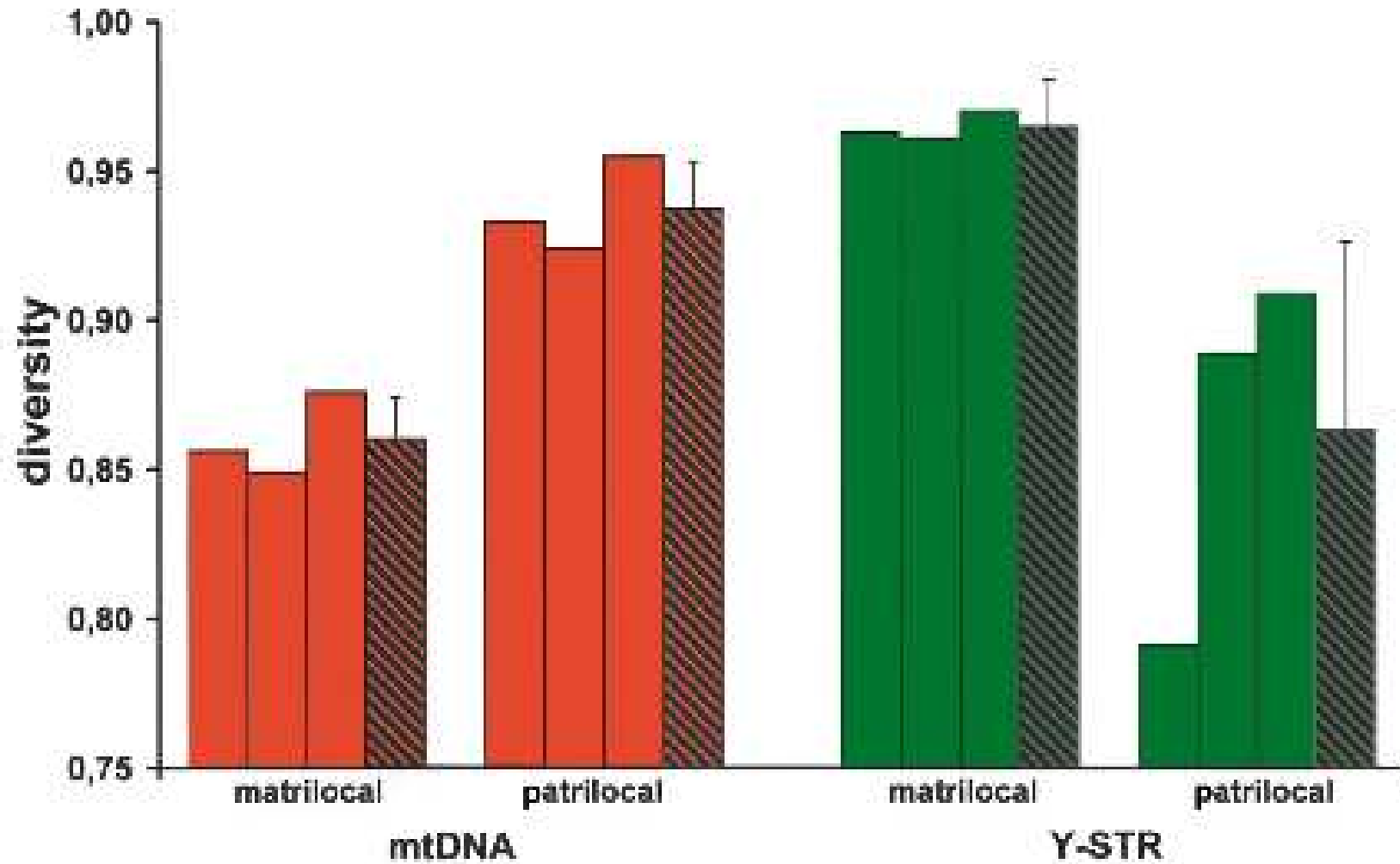
**Graduate School of Natural Sciences, Nagoya City University**

**18 June 2009**

**Stochastic Population Dynamics and Applications in Spatial Ecology**

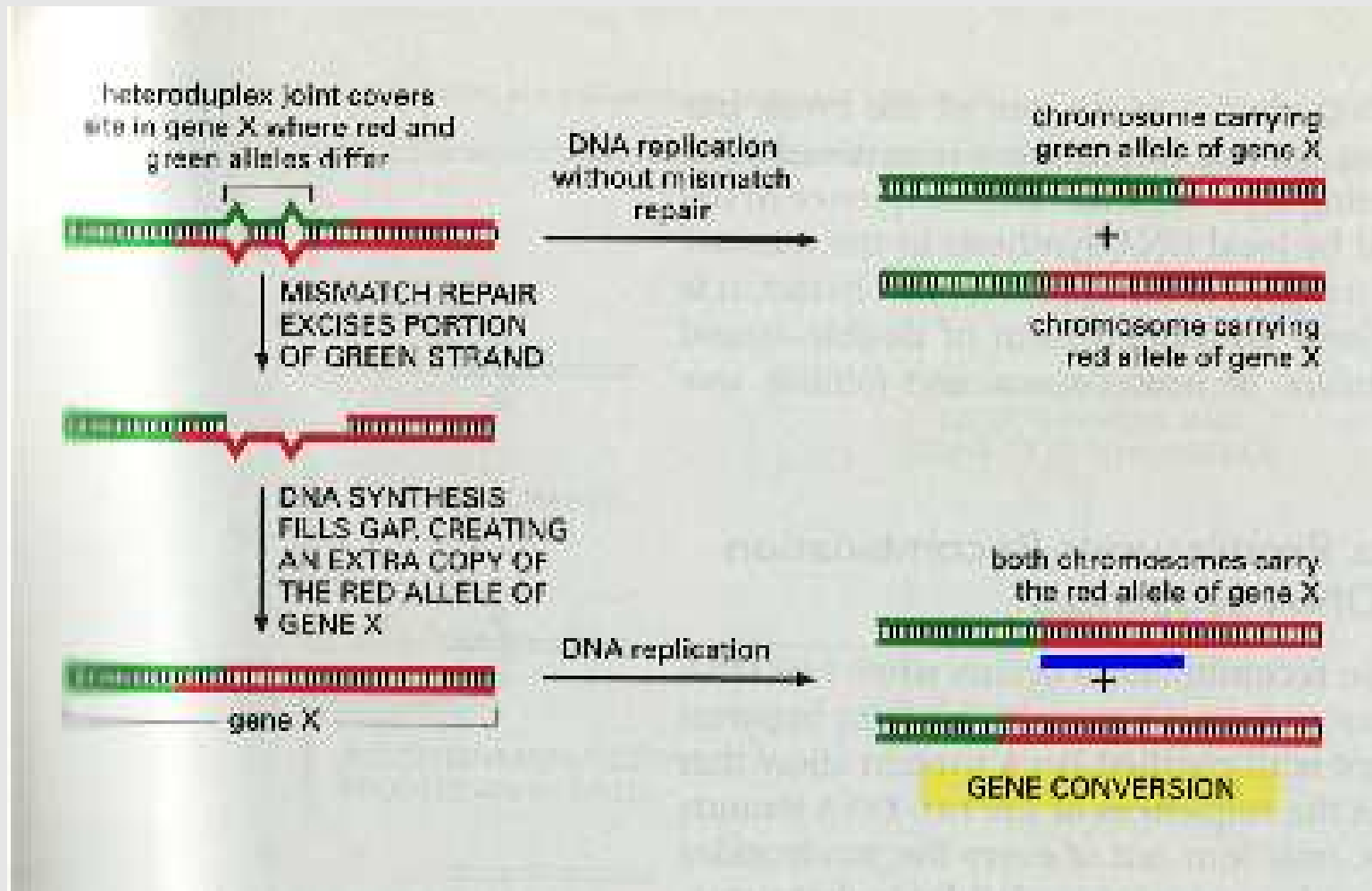
**ICMS, Edinburgh, UK**

# Bias in Human Migration



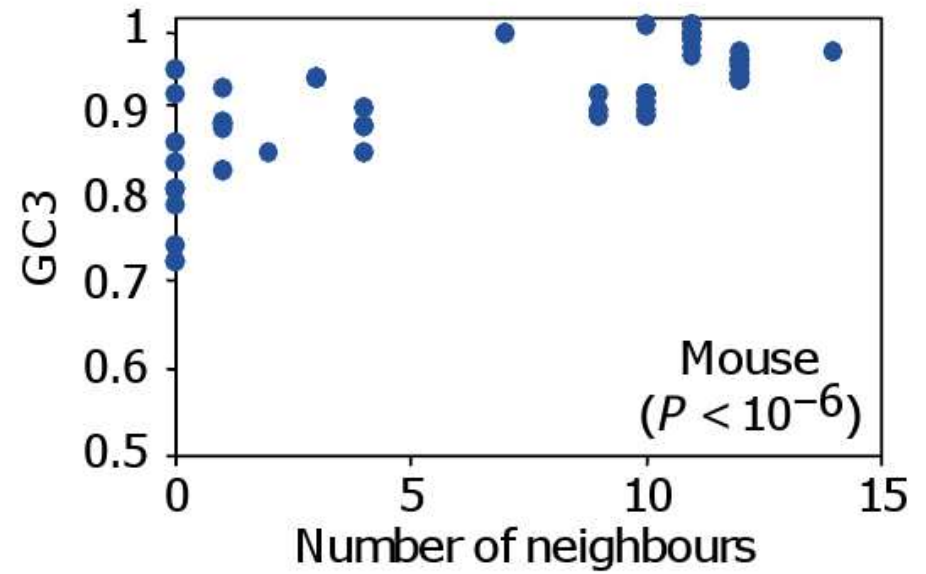
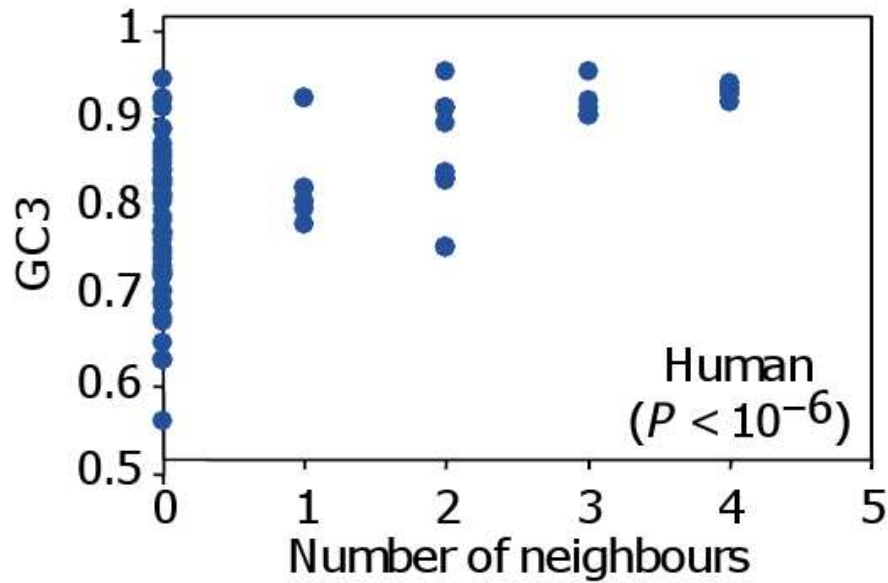
Oota et al. (2001) Nat Genet 29: 20.

# Bias in Gene Conversion at Meiosis



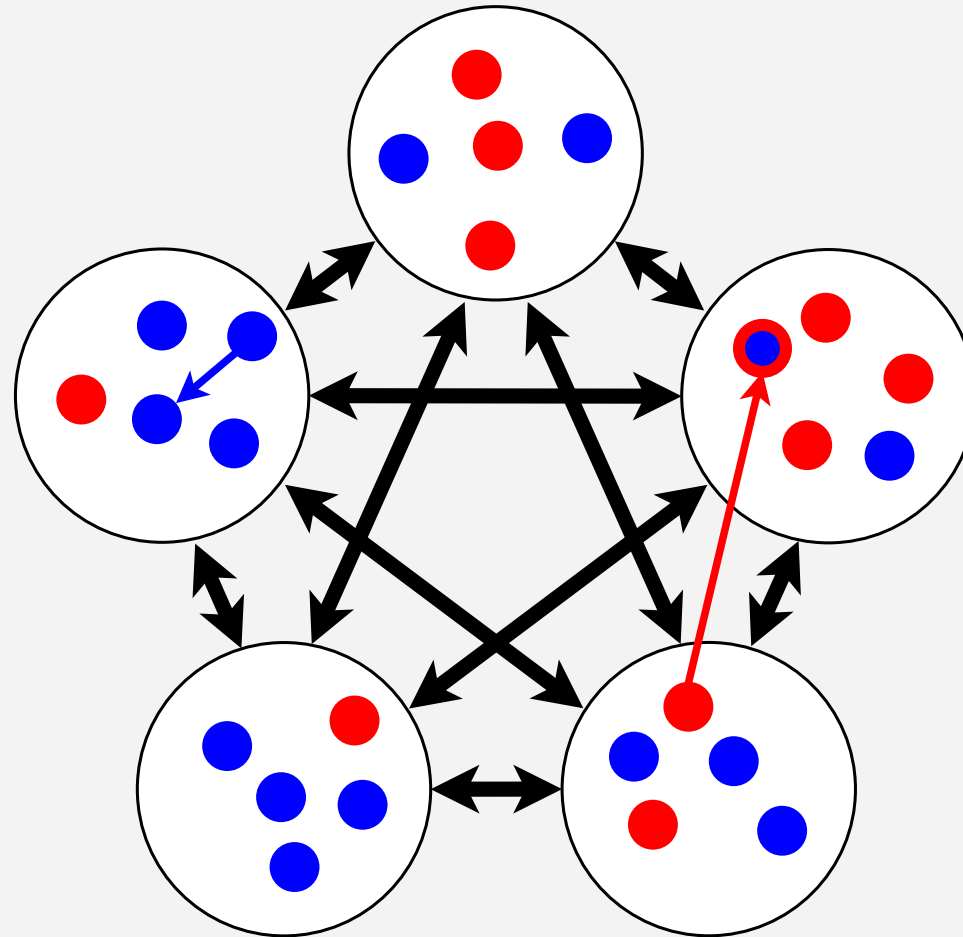
Molecular biology of the cell 4th Eds. (2002)

# GC Content in Multigene Family



Galtier (2003) Trends in Genet 19: 65

# Wright's Island Model with Migration Bias



Reproduction:  $\lambda_0$ , Migration:  $\lambda_0 \xi_i$  for  $A_i$ ,  $i = 1, 2$ .  
 $\xi_2, \xi_1 = m(1 \pm b)/(n - 1)$ ,  $(0 \leq b < 1)$ .

## Moran Model for Migration Bias

A population consists of  $n$  demes, each occupied by  $N$  haploid indiv. The state of the population is a continuous-time Markov chain  $Z(t) = (Z_i(t))$ , where  $Z_i(t)$  is the number of indiv. of type  $A_1$  in the  $i$ -th deme.

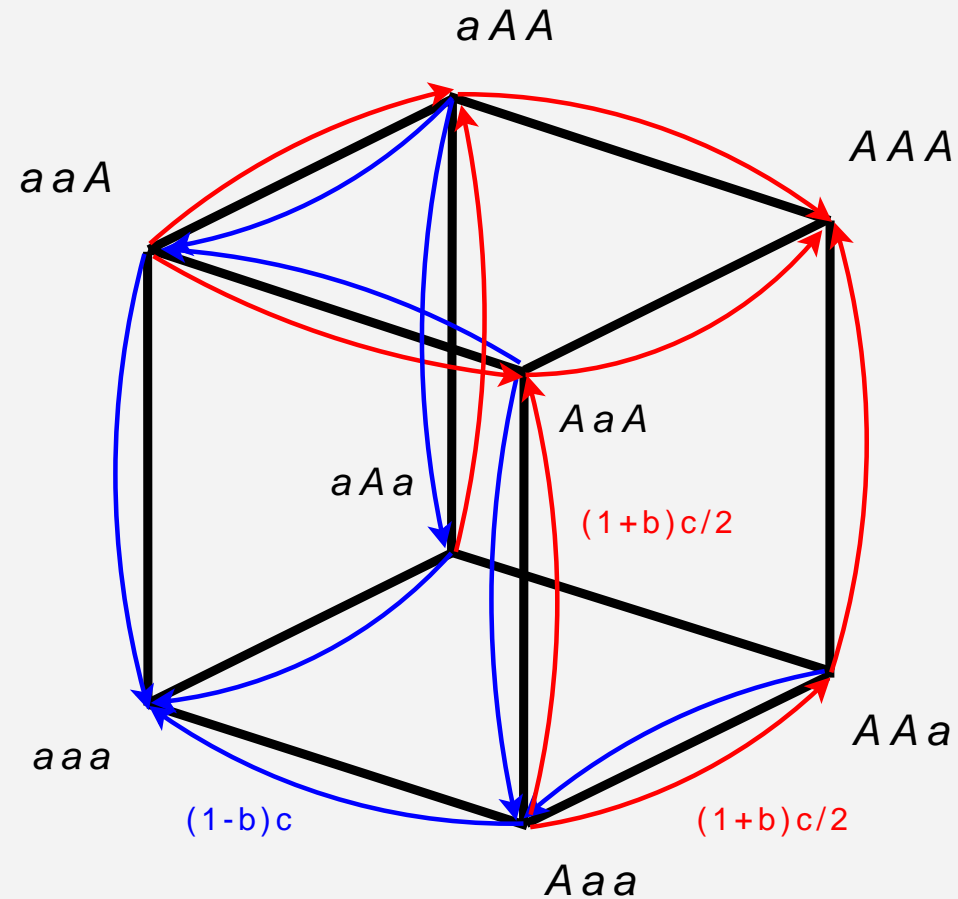
Transition to  $z + e_i$ , is at a rate

$$\lambda_0(z_i + \xi_1 \sum_{k(\neq i)} z_k) \frac{N - z_i}{N} (1 - u) \\ + \lambda_0 \left\{ (N - z_i) + \xi_2 \sum_{k(\neq i)} (N - z_k) \right\} \frac{N - z_i}{N} u.$$

and to  $z - e_i$  is at a rate

$$\lambda_0 \left\{ (N - z_i) + \xi_2 \sum_{k(\neq i)} (N - z_k) \right\} \frac{z_i}{N} (1 - u) + \lambda_0 (z_i + \xi_1 \sum_{k(\neq i)} z_k) \frac{z_i}{N} u.$$

# Gene Conversion among Multigene Family



Reproduction:  $\lambda_0$ , Conversion:  $\lambda_0 \xi_i$  for  $A_i$ ,  $i = 1, 2$ .  
 $\xi_2, \xi_1 = c(1 \pm b)/(n - 1)$ ,  $(0 \leq b < 1)$ .

## Moran Model for Conversion Bias

A population consists of  $2N$  monoecious diploid indivs. The state of the population is a continuous-time Markov chain  $W(t) = (W_\alpha(t))$ , where  $W_\alpha(t)$  is the number of unlinked haplotypes of type  $\alpha$ .

Transition to  $w + e_\alpha$  is at a rate

$$\lambda_0 w_\alpha \frac{2N - w_\alpha}{2N} \left\{ 1 - \sum_{\beta \neq \alpha} Q_{\alpha\beta} \right\} + \lambda_0 \sum_{\beta \neq \alpha} w_\beta \frac{2N - w_\alpha}{2N} Q_{\beta\alpha}$$

and to  $w - e_\alpha$  is at a rate

$$\lambda_0 w_\alpha \frac{w_\alpha}{2N} \sum_{\beta \neq \alpha} Q_{\alpha\beta} + \lambda_0 \sum_{\beta \neq \alpha} w_\beta \frac{w_\alpha}{2N} (1 - Q_{\beta\alpha}),$$

where

$$\frac{W_\alpha(t)}{2N} = \prod_{i=1}^n (X_i(t))^{\alpha_i} (1 - X_i(t))^{1 - \alpha_i}.$$

## Diffusion Limit

Set  $\lambda_0 = N/2$ .  $Nu \rightarrow \theta$ ,  $Nm$  ( $2Nc$ )  $\rightarrow \gamma$ , as  $N \rightarrow \infty$ . Let fraction of  $A_1$  in the  $i$ -th deme at time  $t$  be  $X_i(t)$ .

The diffusion  $\{X(t); t \geq 0\}$  in  $[0, 1]^n$  has a generator:

$$L = L_0 - bL_1,$$

where

$$L_0 = \sum_{i=1}^n \frac{x_i(1-x_i)}{2} \frac{\partial^2}{\partial x_i^2} + \frac{n\gamma}{2(n-1)} \sum_{i=1}^n (\bar{x} - x_i) \frac{\partial}{\partial x_i} + \frac{\theta}{2} \sum_{i=1}^n (1-2x_i) \frac{\partial}{\partial x_i},$$

$$L_1 = \frac{\gamma}{2(n-1)} \sum_{i=1}^n \left\{ (1-2x_i) \sum_{j(\neq i)} x_j + (n-1)x_i \right\} \frac{\partial}{\partial x_i}.$$

Accidentally, both of models have the identical limiting process.

# Probability of Fixation

The probability of fixation of allele  $A_1$  whose initial freqs.  $p$ , satisfies a PDE:

$$L\pi(p) = 0$$

with  $\theta = 0$  and boundary conditions:

$$\pi(0) = 0, \pi(1) = 1, \pi|_{\partial[0,1]^n - \{0,1\}} = \text{finite.}$$

The solution is expanded as

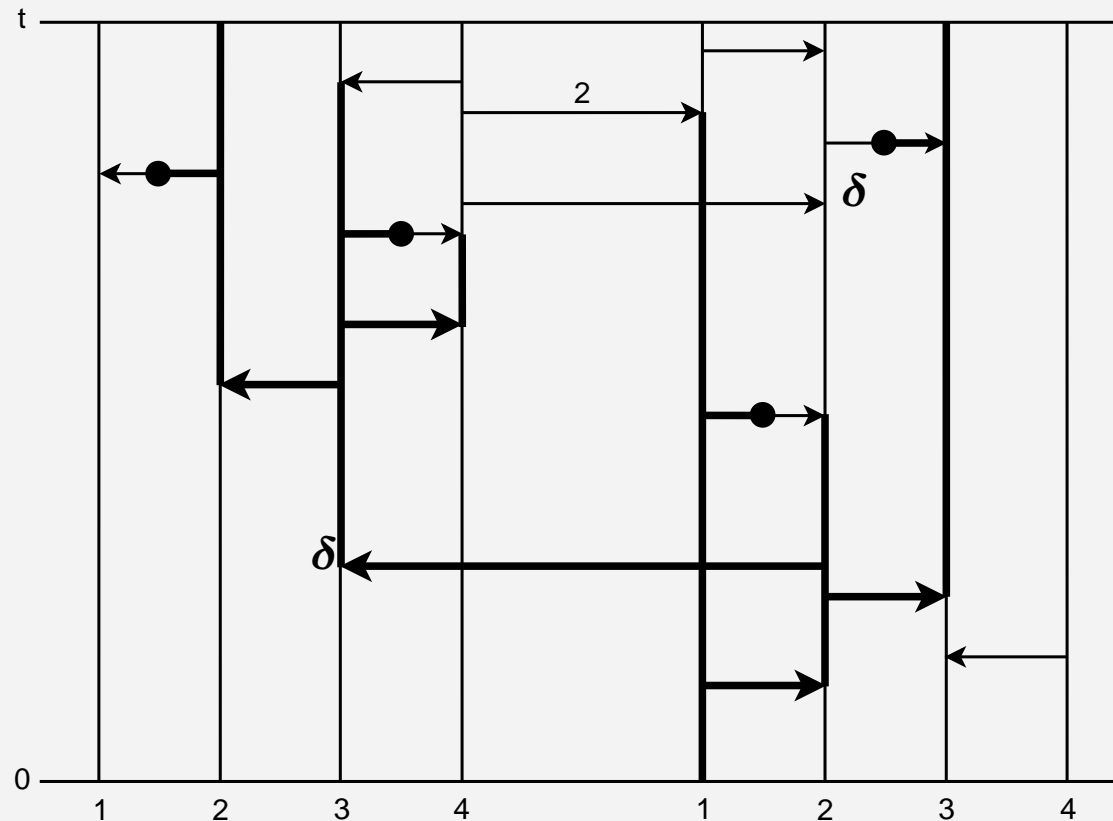
$$\pi(p) = \bar{p} - b \left[ \bar{p}\{n - 1 + n\gamma(1 - \bar{p})\} + \frac{2}{n} \sum_{i < j} p_i p_j \right] + O(b^2).$$

## Biased Voter Model

The Moran model under biased migration has an formulation in terms of BVM. Let  $I = (I_i)$ ,  $I_i = \{1, 2, \dots, N\}$  denote sets of sites, where  $I_i$  is the set of sites in the  $i$ ,  $i = 1, 2, \dots, n$ -th graph. The continuous-time Markov process  $\{\eta_t; t \geq 0\}$ ,  $\eta_t : I \rightarrow \{1, 2\}$ , evolves according to the following rules:

- i. The indiv. at  $x \in I_i$  produces an offs. at rate  $\lambda_0$  within  $I_i$ .
- ii. The offs. is of the same type as the parent with prob.  $1 - u$  and of the other type with prob.  $u$ .
- iii. The indiv. at  $x \in I_i$  produces offs. in  $I_j$  at rates depending on the type. If  $\eta_t(x) = 1$  (2), then the rate is  $\lambda_0 \xi_1$  ( $\lambda_0 \xi_2$ ).  $\xi_2 - \xi_1 = 2mb/(n - 1)$ .
- iv. At the time of the birth event, one of the  $N$  site is chosen at random and the individual at the site is replaced by the offspring.

# Percolation Diagram



$W_i^{x,y} \sim Po(\lambda_0/N)$ ,  $Z_{i,j}^{x,y} \sim Po(\lambda_0\xi_2/N)$ ,  $U_{i,j}^{x,y} \sim Uni(0, 1)$ . At  $W_i^{x,y}$ , an arrow from  $x \in I_i$  to  $y \in I_i$ . At  $Z_{i,j}^{x,y}$ , an arrow from  $x \in I_i$  to  $y \in I_j$ . If  $U_{i,j}^{x,y} < \xi_1/\xi_2$ , we put  $\delta$ , otherwise we put 2. We have  $\delta$ -arrows entering at rate  $\lambda_0\xi_1$  and 2-arrows at rate  $\lambda_0(\xi_2 - \xi_1)$ .



# Ancestral Bias Graph

Assume there are  $k = (k_i), i = 1, 2, \dots, n$ , particles in the dual process.

- Coalescing event

$$\lambda_0 k_i \frac{k_i - 1}{N} = \frac{k_i(k_i - 1)}{2}.$$

- Migration event

$$\lambda_0 \xi_1 k_i (N - k_j) \rightarrow \frac{(1 - b)\gamma}{2(n - 1)} k_i.$$

- Branching event

$$\lambda_0 (\xi_2 - \xi_1) k_i (N - k_j) \rightarrow \frac{b\gamma}{n - 1} k_i.$$

We call the random graph generated by the dual process in the limit  $N \rightarrow \infty$  as the Ancestral Bias Graph.

## Stationary Measure of Size Process

Let  $A_n(t) = (A_{n,i}(t))$  denote the number of particles that are present in ABG at the dual time  $t$ , where  $A_{n,i}(t)$  is the number of particles contained in  $I_i$ .

The size process  $A_n(t)$  is an  $n$ -dimensional birth and death process. Thanks to the moment duality between  $A_n(t)$  and the Wright-Fisher diffusion, the stationary measure of  $A_n(t)$  can be obtained by the probability of fixation in the Wright-Fisher diffusion (Mano, 2009, TPB 75: 164).

$$\phi(e_i) = \frac{1}{n} - b \left( \frac{n-1}{n} - \gamma \right) + O(b^2),$$

$$\phi(2e_i) = b \frac{\gamma}{n} + O(b^2),$$

$$\phi(e_i + e_j) = b \frac{2}{n} (1 + \gamma) + O(b^2).$$

For other configurations,  $O(b^2)$ .

# Sampling Distribution

A scheme using MCMC to simulate along sample path of ABG can approximate the sampling distribution in stationarity via a recursion. For

$$\begin{aligned}
 q(a, d) &= \prod_{i=1}^n \frac{(f_i)!}{a_i! d_i!} \mathbb{E}_f [X_i^{a_i} (1 - X_i)^{d_i}], \\
 r(a, d)q(a, d) &= \sum_{i=1}^n \{ (a_i - 1) f_i q(a - e_i, d) + (d_i - 1) f_i q(a, d - e_i) \} \\
 &+ \theta \sum_{i=1}^n \{ (d_i + 1) q(a - e_i, d + e_i) + (a_i + 1) q(a + e_i, d - e_i) \} \\
 &+ \frac{(1 - b)\gamma}{n - 1} \sum_{j \neq i} f_i \left\{ \frac{a_j + 1}{f_j + 1} q(a - e_i + e_j, d) + \frac{d_j + 1}{f_j + 1} q(a, d - e_i + e_j) \right\} \\
 &+ \frac{2b\gamma}{n - 1} \sum_{i \neq j} \left\{ \frac{d_i(d_j + 1)}{f_j + 1} q(a, d + e_j) + \frac{f_i(a_j + 1)}{f_j + 1} q(a + e_j, d) \right. \\
 &\left. + \frac{(a_i + 1)(d_j + 1)}{f_j + 1} q(a + e_i, d - e_i + e_j) \right\}.
 \end{aligned}$$

## Sample of Size Two

**Especially,**

$$q(2e_i, 0) = \frac{\gamma'/n + \theta + (1 + 2\theta + \gamma')(\theta - b\gamma)}{2\{\gamma'/n + 2\theta(1 + 2\theta + \gamma')\}} + O(b^2),$$

$$q(e_i + e_j, 0) = \frac{\gamma'/n + (1 + 2\theta + \gamma')(\theta - b\gamma)}{2\{\gamma'/n + 2\theta(1 + 2\theta + \gamma')\}} + O(b^2),$$

$$q(e_i, e_i) = \frac{\theta(2\theta + \gamma')}{\gamma'/n + 2\theta(1 + 2\theta + \gamma')} + O(b^2),$$

$$q(e_i, e_j) = \frac{\theta(1 + 2\theta + n\gamma')}{2\{\gamma'/n + 2\theta(1 + 2\theta + \gamma')\}} + O(b^2).$$

$q(0, 2e_i), q(0, e_i + e_j)$  are given by  $q(2e_i, 0), q(e_i + e_j, 0)$ , respectively, with replacing  $b$  by  $(-b)$ .  $\gamma' = n\gamma/(n - 1)$ .

## Strong Migration Limit

Nagylaki (1980, JMB 9: 101) established the strong-migration limit for geographically structured population.

Let  $X(t) = \sum_{i=1}^n Z_i(t)/(nN)$ . The limiting ( $N \rightarrow \infty$ ) diffusion of the continuous-time Markov chain  $X(t)$  with  $m$  and  $n$  fixed has a generator

$$L = \frac{x(1-x)}{2} \frac{\partial^2}{\partial x^2} - \left\{ bn\gamma x(1-x) - \frac{n\theta}{2}(1-2x) \right\} \frac{\partial}{\partial x}$$

is identical with that for selection, which has been extensively investigated, where the selection intensity is  $2bnm$ . In the strong migration limit, ABG reduces to the ancestral selection graph introduced by Krone and Neuhauser (1997, TPB 51: 210).

## Bias in Mouse Histone Gene Family

Mouse histone H2A gene family (operationary defined) contains 36 sequences, whose GC3 88.0% (average GC3 of the mouse genome is 42%). This high GC probably comes from BGC.

Moment estimate of conversion rate via the sampling formula,  $c = 1.26 \times 10^{-7}$ . According to the formula of prob. of fixation, expected GC proportion is

$$\eta(t) = \tilde{\eta} + (\eta(0) - \tilde{\eta})e^{-ut}, \quad \tilde{\eta} = \frac{1 - (n - 1)(1 + \gamma')b}{2}.$$

If the GC content reaches its equilibrium,  $\tilde{\eta} = 0.880$  and we have  $b = 0.0209$ .

Recent estimate of allelic conversion bias in high recombination region of human genome is approximately 4%.

# Summary

**Bias in migration and that in gene conversion are mathematically equivalent phenomenon, as long as we consider the diffusive limits.**

**We demonstrated that few percent BGC among multigene family could cause substantial increase of the GC content in a multigene family.**

**It is reasonable to assume that there might be slight bias among migration rates of alleles. It seems likely that slight migration bias also have large impact on population differentiation and speciation in natural populations.**