

6. 構造を持つ合祖過程

Kingman の合祖理論は可換モデル、すなわち個体を入れ替えても子供の数の分布はかわらないという条件のもとで得られる。しかし、集団に地理的な構造がある場合、あるいは対立遺伝子間で自然選択が存在する場合にはこの可換性が成り立たない。6. 1 節～6. 4 節では集団が幾つかの分集団に分かれて繁殖する地理的構造を考慮に入れた遺伝子系図について紹介する。また6. 5 節では、自然選択が存在する場合、対立遺伝子タイプによる構造が現れる遺伝子系図モデルについて紹介する。

6. 1 地理的構造を持つ集団における遺伝子系図

生物集団は全集団が一つの任意交配集団と見なされることは稀であり、地理的な隔たりが交配のチャンスについて制限を与え、その生物種の遺伝的多様性や進化に大きな影響を与える。遺伝的変異が少なく生物種としては比較的若いと考えられる人類集団においても様々な地理的変異が見られる。ここでは、このような地理的構造を考慮に入れた遺伝子系図のモデルを紹介する。

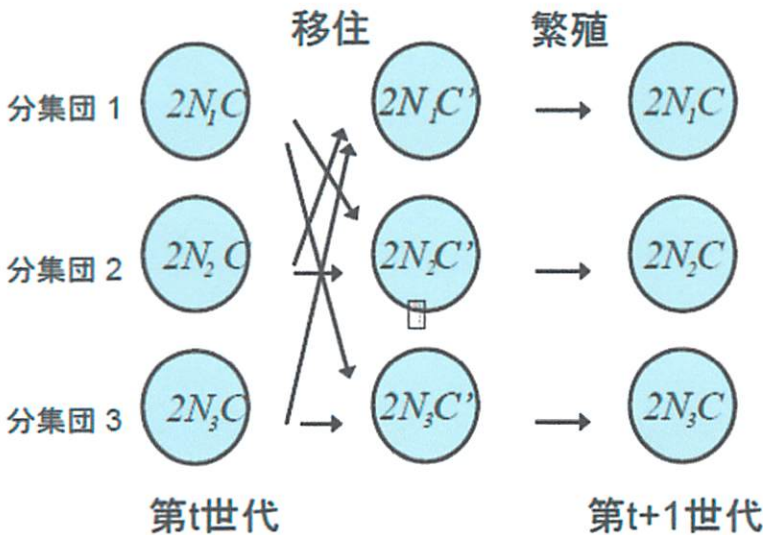
有限個の小集団に分割された半数体生物集団を仮定しよう。各分集団を k, j などで表し

$K = \{k, j, \dots\}$ を全分集団の集合とする。 $|K|$ で分集団の総数を表し、ここでは K は有限集

合 ($|K| < \infty$) とする。分集団 $k \in K$ の遺伝子の個数を $2N_k C$ (一定) とする。親個体は独立

に分集団間を移住し、移住先の分集団ごとに配偶子の遺伝子プールを作る。その後、各分集団で独立に Wright-Fisher モデルを含む可換モデルに従って、それぞれの分集団 $k \in K$ の $2N_k C$ の個体 (遺伝子) からなる次世代の集団を作る。繁殖と移住で1世代と考えることにする。

分集団構造をもつモデル



1 世代 1 個体当たり分集団 $k \in K$ から $j \in K$ への移住率 (移住確率) を $\frac{q_{k,j}}{C}$ (ただし $k \neq j$)、および $k = j$ のとき $1 + \frac{q_{k,k}}{C}$ (ただし $q_{k,k} = -\sum_{j(\neq k)} q_{k,j}$) とする。また、行列 Q を $Q = (q_{k,j}; k, j \in K)$ とする。このとき時間を逆に見て、分集団 $k \in K$ にいる個体が 1 世代前に $j \in K$ から移住してきた確率を $P(j|k)$ とすると次の補助定理を得る。

補助定理 6. 1

$$P(j|k) = \delta_{k,j} + \frac{m_{k,j}}{C} + O\left(\frac{1}{C^2}\right) \quad k, j \in K$$

ただし $M = (m_{k,j}; k, j \in K)$ は後ろ向き移住率行列と呼ばれ次式で与えられる。

$$m_{k,j} = \begin{cases} \frac{N_j}{N_k} q_{j,k} & \text{if } k \neq j \\ -\frac{1}{N_k} \left(\sum_{\substack{i \in K \\ i \neq k}} N_i q_{i,k} \right) & \text{if } k = j \end{cases} \quad (6.1)$$

(証明)

分集団間を上で定義した移住率 $Q = (q_{j,k})$ で毎世代動く個体の時刻 t の位置を Z_t とする。1 世代を $\Delta t = \frac{1}{C}$ と見なす時間スケールを取ると、全集団からランダムに選んだある個体の 1 世代当たりの移動より

$$P(j|k) = P(Z_0 = j | Z_{\Delta t} = k) = \frac{P(Z_0 = j, Z_{\Delta t} = k)}{P(Z_{\Delta t} = k)} = \frac{P(Z_0 = j)P(Z_{\Delta t} = k | Z_0 = j)}{\sum_{i \in K} P(Z_0 = i)P(Z_{\Delta t} = k | Z_0 = i)}$$

全集団からランダムに選んだ個体なので、 $P(Z_0 = i) = \frac{2N_i C}{2NC} = \frac{N_i}{N}$ 、ただし $N = \sum_{i \in K} N_i$ 。

よって $k \neq j$ のとき

$$\begin{aligned} P(j|k) &= \left(\frac{N_j}{N} \times \frac{q_{j,k}}{C} \right) \bigg/ \left(\sum_{\substack{i \in K \\ i \neq k}} \frac{N_i}{N} \times \frac{q_{i,k}}{C} + \frac{N_k}{N} \times \left(1 + \frac{q_{k,k}}{C} \right) \right) = \frac{N_j q_{j,k}}{\sum_{i \in K} N_i q_{i,k} + N_k C} \\ &= \frac{N_j q_{j,k}}{N_k} \times \frac{1}{C} + O\left(\frac{1}{C^2}\right) \end{aligned}$$

同様に $k = j$ のとき、

$$\begin{aligned}
P(k|k) &= \left(\frac{N_k}{N} \times \left(1 + \frac{q_{k,k}}{C} \right) \right) / \left(\sum_{\substack{i \in K \\ i \neq k}} \frac{N_i}{N} \times \frac{q_{i,k}}{C} + \frac{N_k}{N} \times \left(1 + \frac{q_{k,k}}{C} \right) \right) = \frac{N_k(C + q_{k,k})}{\sum_{i \in K} N_i q_{i,k} + N_k C} \\
&= 1 - \sum_{\substack{i \in K \\ i \neq k}} \frac{N_i q_{i,k}}{N_k C} + O\left(\frac{1}{C^2}\right)
\end{aligned}$$

以上より(6.1)が成り立つ。

$\sum_{j \in K} m_{k,j} = 0$ が成り立つので、 $C \rightarrow \infty$ ($\Delta t \rightarrow 0$) のとき、 $|K| \times |K|$ の行列

$M = (m_{k,j}; k, j \in K)$ は分集団の集合 K 上に連続時間マルコフ連鎖を生成する。推移確率を

$P_t(k, j) = P(Z_t = j | Z_0 = k)$ とすると、コルモゴロフ方程式

$\frac{d}{dt} P_t(k, j) = \sum_{i \in K} m_{k,i} P_t(i, j) = \sum_{i \in K} P_t(k, i) m_{i,j}$ が成り立つ。推移確率行列を

$P_t = (P_t(k, j); k, j \in K)$ とすると、 $P_t = \exp(tM)$ で表わされる。このマルコフ連鎖は既約と仮定する。次に複数の個体 (遺伝子) の移住について考えてみよう。複数の個体の位置を多重指数 $\alpha = (\alpha_k; k \in K)$ で表わす。 α_k は着目している複数の個体の中で分集団 $k \in K$ に位置する個体の数を表わす。従って $|\alpha| = \sum_{k \in K} \alpha_k$ は全個体数である。このような多重指数

α 全体の集合を $I = \{\alpha = (\alpha_k; k \in K); |\alpha| = \sum_{k \in K} \alpha_k < \infty\}$ とする。各個体は行列

$Q = (q_{k,j}; k, j \in K)$ と前に定義した移住率に従って毎世代独立に分集団間を移住すると仮定する。現在の配置が $\alpha = (\alpha_k; k \in K)$ であるとき、1世代前の配置が $\beta = (\beta_k; k \in K)$ である確率を $P_{\Delta t}(\alpha, \beta)$ とする。複数の個体が1世代で異なる分集団へ移住する確率は $O\left(\frac{1}{C^2}\right)$ のオーダーの無限小となる。よって

$$P_{\Delta t}(\alpha, \beta) = \begin{cases} \alpha_k m_{k,j} \Delta t + O((\Delta t)^2) & (\beta = \alpha - \varepsilon^k + \varepsilon^j \ (k \neq j) \text{ のとき}) \\ 1 + \sum_{k \in K} \alpha_k m_{k,k} \Delta t + O((\Delta t)^2) & (\beta = \alpha \text{ のとき}) \\ O((\Delta t)^2) & \text{その他} \end{cases} \quad (6.2)$$

ここで $\Delta t = \frac{1}{C}$ 、 $\varepsilon^k \in I$ は $\varepsilon^k_j = \delta_{k,j}$ で定義される単位ベクトルとする。

$C \rightarrow \infty$ の連続時間近似を行うと $|\alpha|$ 個体の時刻 t での配置を $\alpha(t)$ とすると、移住による推移

確率 $P_t(\alpha, \beta) = P(\alpha(t) = \beta | \alpha(0) = \alpha)$ は次のコルモゴロフ後ろ向き方程式を満たす。

$$\frac{d}{dt} P_t(\alpha, \gamma) = \sum_{k,j} \alpha_k m_{k,j} P_t(\alpha - \varepsilon^k + \varepsilon^j, \gamma)$$

集団の様々な場所から遺伝子をサンプルし、これらの遺伝子の祖先あるいは系図を考察する。そこで分集団 $k \in K$ から α_k 個の遺伝子をサンプルし、その全体をベクトル

$\alpha = (\alpha_k; k \in K)$ で表す。 $|\alpha| = \sum_{k \in K} \alpha_k$ はサンプル遺伝子の総数である。このサンプル遺伝

子 $\alpha = (\alpha_k; k \in K)$ の祖先をたどりその遺伝子系図を表現するマルコフ過程を求める。

$\alpha(t) = (\alpha_k(t); k \in K)$ を $[Ct]$ 世代遡った祖先集団におけるサンプル α の祖先遺伝子の空間配置とする。すなわち $\alpha_k(t)$ は分集団 k に見出されるサンプル α の祖先遺伝子の数で

ある。 $|\alpha(t)| = \sum_{k \in K} \alpha_k(t)$ は $[Ct]$ 世代前の α の祖先遺伝子の総数である。このとき、

$C \rightarrow \infty$ の連続時間近似を行うと次のコルモゴロフ後退方程式で表現される遺伝子の系図を表すマルコフ過程を得る。

定理 6. 2

$P_t(\gamma | \alpha) = P(\alpha(t) = \gamma | \alpha(0) = \alpha)$ を初期状態 $\alpha(0) = \alpha$ から $\alpha(t) = \gamma$ への推移確率

とする。このとき次の方程式が成り立つ(Notohara(1990))。

$$\frac{d}{dt} P_t(\gamma | \alpha) = \sum_{\beta \in I} R_{\alpha, \beta} P_t(\gamma | \beta) \quad P_0(\gamma | \alpha) = \delta_{\alpha, \gamma}$$

$$R_{\alpha, \beta} = \begin{cases} \alpha_k m_{k,j} & \text{if } \beta = \alpha - \varepsilon^k + \varepsilon^j \\ \frac{\alpha_k(\alpha_k - 1)}{4N_k} & \text{if } \beta = \alpha - \varepsilon^k \\ -\left\{ \sum_{k \in S} \frac{\alpha_k(\alpha_k - 1)}{4N_k} + \sum_{k \in S} \alpha_k |m_{k,k}| \right\} & \text{if } \beta = \alpha \\ 0 & \text{otherwise} \end{cases} \quad (6.3)$$

ここで $M = (m_{k,j}; k, j \in K)$ は後ろ向き移住率、 $R = (R_{\alpha, \beta}; \alpha, \beta \in I)$ とする。

明らかに $\sum_{\beta \in I} R_{\alpha, \beta} = 0$ であり(6.3)により I 上に保存的なマルコフ過程が定義される。これが

移住と合祖を含む地理的構造を持つ遺伝子の系図過程である。この遺伝子系図過程を

α_i ($\alpha_0 = \alpha$)と書くことにする。

連続時間マルコフ連鎖の性質より状態 α での滞在時間を $\tau(\alpha)$ とすると $\tau(\alpha)$ はパラメータ $|R_{\alpha,\alpha}|$ の指数分布に従う。

$$P(\tau(\alpha) > t) = \exp[R_{\alpha,\alpha}t] = \exp\left[-\left\{\sum_{k \in S} \frac{\alpha_k(\alpha_k - 1)}{4N_k} + \sum_{k \in S} \alpha_k |m_{k,k}|\right\}t\right] \quad (6.4)$$

この滞在時間の後、状態 α から β ($\beta \neq \alpha$) へのジャンプが確率 $\frac{R_{\alpha,\beta}}{|R_{\alpha,\alpha}|}$ で起こる。

方程式(6.2)(6.3)は時間を遡って見た時、各個体の移住の独立性と各分集団内での繁殖の独立性より1世代内で後ろ向きの移住と合祖が同時に起こる確率、複数の分集団で同時に合祖が起こる確率は $O\left(\frac{1}{C^2}\right)$ のオーダーの無限小ということから導かれる。

ただし、(6.3)式の導出は数学的には厳密なものではない。なぜなら、後ろ向きの移住によって1世代前の個体の配置を考えると、各分集団の個体数は各分集団の定まった個体数 $2N_k C$ に戻っていなければならない。従って今注目しているサンプル遺伝子の後ろ向きの移住は独立ではない。しかし、1世代あたり2個体以上が移住する確率は $O\left(\frac{1}{C^2}\right)$ のオーダーの無限小となり、複数の個体の移住が無視できるということより(6.2)が導かれると考えられる。また数学的には無限個の分集団を含むモデルに拡張することもできる。保存的な移住という条件の下で、Herbots(1997)はここで述べたものとは異なるモデル設定で(6.3)式を厳密に導いている。ここで保存的移住とは移住の前後で個体数の変化がないという意味で等式 $N_i \sum_{j \neq i} q_{ij} = \sum_{j \neq i} N_j q_{ji}$ が成り立つときに言う。さらに、Kozakai, Shimizu, and

Notohara(2016)は Herbots(1997)のモデルを非保存的移住まで拡張し、(6.3)が成り立つことを示した。これは以下の様な移住、繁殖から成る離散時間モデルである。 $K = \{i, j, k, \dots\}$ を分集団の集合とする。分集団 $i \in K$ の個体数を $N_i = 2c_i N$ とする。 c_i は分集団の相対的大きさを表すパラメーターである。 $Q = (q_{k,j})_{k,j \in K}$, $\sum_{j(\neq k)} q_{k,j} \leq 1$ を前向きの移住率とする。

ただし、移住は各個体が独立ではなく、各分集団 $i \in K$ において $N_i q_{i,j} = 2c_i N q_{i,j}$ 個の個体が分集団 $i \in K$ から $j \in K, (j \neq i)$ へ移住するものとする。 $q_{i,j}$ は分集団 $i \in K$ の個体の中で、分集団 $j \in K$ へ移住する個体の割合である。各分集団でランダムに個体をあらかじめ移住先によって配分しておくモデル設定である。 $\{c_i; i \in K\}$ 及び $(q_{i,j})$ について次の仮定をする。

(1) 全ての $i \in K$ について、 $1 \leq c_i < K$ (定数)

$$(2) \quad q_{i,j} = \frac{q_{i,j}^*}{4N}, i \neq j; \sup_i \left(\sum_{j(i \neq j)} q_{i,j}^* \right) < \infty, \sup_j \left(\sum_{i(i \neq j)} q_{i,j}^* \right) < \infty$$

このとき、移住後の分集団 $i \in K$ のサイズ N_i^* は

$$N_i^* = \sum_{k(i \neq j)} N_k q_{k,i} + N_i \left(1 - \sum_{j(i \neq j)} q_{i,j} \right) = 2c_i N + \frac{Q_i}{2} \text{ と表される。}$$

ただし $Q_i = \sum_{k(i \neq j)} c_k q_{k,i}^* - c_i \sum_{j(i \neq j)} q_{i,j}^*$ 、 $\frac{Q_i}{2}$ は分集団 $i \in K$ における移住の前後で増減した個体数を表す。全ての $i \in K$ について $Q_i = 0$ のとき、保存的移住と言う (Herbots(1997))。

後ろ向き移住率 $M = (m_{i,j})$ は $m_{i,j} = \frac{N_j q_{j,i}}{N_i^*}$, ($i \neq j$) で与えられる。 $m_{i,j}$ は分集団 $i \in K$ の個体の中で、分集団 $j \in K$ から移住してきた個体数の割合を表す。この移住の後、各分集団で独立に可換モデルによる繁殖で元の個体数に戻る。以上のモデル設定で N 世代を単位時間に取り、 $N \rightarrow \infty$ の極限を取ると定理 6. 2、(6.3)式 (ただし、 N_k は全て c_k に置き換える) で表現される連続時間のマルコフ連鎖に弱収束することが示される (Kozakai, Shimizu and Notohara(2016))。

この様な生物集団が幾つかの分集団に分割されたモデルは飛び石モデル (stepping stone model) と呼ばれ kimura(1953)によって提案された。その後 Kimura&Weiss(1964)、Weiss&Kimura(1965), Malecot(1967,1975)、, Sawyer(1976b)、Maruyama(1970a,b,c,d、等) Nagylaki(1974a,c、1976,1977、等)など多くの研究がある。(6.2)(6.3)で定義される地理的構造を持つ合祖過程 (Structured Coalescent Model) は任意サイズのサンプル遺伝子の系図という視点から飛び石モデルの自然な拡張と考えられる。

K を可算無限集合としよう。系図過程 α_t ($\alpha_0 = \alpha$) は次の飛び石モデル (stepping stone model) を表現する無限次元拡散過程の双対過程として現れる。二つの中立な対立遺伝子 A, a を考え、アレル A の分集団 $k \in K$ での頻度を x_k 、そして分集団 $k \in K$ のサイズパラメーターを N_k とし次の方程式で表現される無限次元拡散過程を考える。

$$\frac{\partial}{\partial t} P_t(X, Y) = \sum_{k \in K} \frac{x_k(1-x_k)}{4N_k} \frac{\partial^2}{\partial x_k^2} P_t(X, Y) + \sum_{k \in S} \left\{ \sum_{j \in S} m_{k,j} x_j \right\} \frac{\partial}{\partial x_k} P_t(X, Y) \quad (6.5)$$

遺伝子頻度 $X = (x_k; k \in K)$ の有限個の頻度の単項式を $X^\alpha = \prod_{k \in K} x_k^{\alpha_k}$ で定義する。ただし

$\alpha = (\alpha_k; k \in K), |\alpha| = \sum_{k \in K} \alpha_k < \infty$ とする。このとき、 $f_t(\alpha) = E[X_t^\alpha]$ とすると半群の性質

$$\text{より } \frac{d}{dt} f_t(\alpha) = \sum_{\beta \in I} R_{\alpha, \beta} f_t(\beta), \quad f_0(\alpha) = X^\alpha \quad (6.6)$$

ここで $(R_{\alpha, \beta}; \alpha, \beta \in I)$ は(6.3)で定義される定数である。 $R = (R_{\alpha, \beta})$ から生成されるマルコフ連鎖は遺伝子の系図過程 $\alpha(t)$ なので、 $E_X[X_t^\alpha] = E_\alpha[X^{\alpha(t)}]$ と表現される。左辺は初期条件 $X_0 = X$ の下での遺伝子頻度の拡散過程 X_t のモーメントであり右辺は初期頻度 X のモーメントの指数として現れる遺伝子系図過程に関する期待値である。この等式は左辺すなわち時刻 t のサンプル α が全てアレル A である確率は、右辺すなわち時間を t 遡った祖先集団においてその祖先が全てアレル A である確率に等しいことを表現している。この関係は双対性と呼ばれ、遺伝子頻度 X_t と遺伝子系図過程 $\alpha(t)$ は双対な確率過程であると言う。この双対性を利用して、遺伝子系図過程 $\alpha(t)$ の解析から拡散過程 X_t の種々の性質を導くことができる。特に拡散過程 X_t の定常分布およびエルゴート性については Shiga(1980a,b)で研究された。

6. 2 合祖時間分布と分離サイトの数の分布

Kingman の合祖過程の場合と同様にサンプル遺伝子が共通な 1 つの祖先に到達するまでの待ち時間を $T = \inf\{t; |\alpha_t| = 1\}$ とする。

定理 6. 3. $\Psi(\alpha) = E[\exp(-\lambda T) | \alpha(0) = \alpha]$ (ただし $\lambda \geq 0$) を T の母関数とする。 $\Psi(\alpha)$

は T の分布のラプラス変換である。このとき $\Psi(\alpha)$ は次式を満たす。

$$\sum_{\beta \in I} R_{\alpha, \beta} \Psi(\beta) = \lambda \Psi(\alpha) \quad \text{、境界条件は全ての } k \in K \text{ に対して } \Psi(\epsilon^k) = 1. \quad (6.7)$$

(証明) 状態 α での滞在時間 $\tau(\alpha)$ とジャンプ確率 $\frac{R_{\alpha, \beta}}{|R_{\alpha, \alpha}|}$ を用いて

$$\begin{aligned} \Psi(\alpha) &= E[\exp(-\lambda T) | \alpha(0) = \alpha] = E[\exp(-\lambda(T - \tau(\alpha)) - \lambda\tau(\alpha)) | \alpha] \\ &= E[E[\exp(-\lambda\tau(\alpha)) \exp(-\lambda(T - \tau(\alpha))) | F_{\tau(\alpha)}] | \alpha] \\ &= E[\exp(-\lambda\tau(\alpha)) E[\exp(-\lambda(T - \tau(\alpha))) | F_{\tau(\alpha)}] | \alpha] \quad (\text{強マルコフ性(付録I参照)}) \\ &= E[\exp(-\lambda\tau(\alpha)) E[\exp(-\lambda T) | \alpha(\tau(\alpha))] | \alpha] \\ &= E[\exp(-\lambda\tau(\alpha)) | \alpha] \sum_{\substack{\beta \in I \\ \beta \neq \alpha}} \frac{R_{\alpha, \beta}}{|R_{\alpha, \alpha}|} \Psi(\beta) = \frac{-R_{\alpha, \alpha}}{\lambda + R_{\alpha, \alpha}} \sum_{\beta \neq \alpha} \frac{R_{\alpha, \beta}}{(-R_{\alpha, \alpha})} \Psi(\beta) \end{aligned}$$

$\alpha = \varepsilon^k$ のとき $T = 0$ より (6.7) 並びに境界条件を得る。

(6.7) を一般に解くことは困難であるが、集団構造が単純でサンプル数が小さな場合については具体的に $\Psi(\alpha)$ を求めることができる。その例を紹介しよう。

<例> : d次元トーラス状格子モデル

分集団 K が d次元のトーラス状の格子空間上に配置され、各分集団のサイズパラメーターが等しく空間的に一様な移住率を持つ場合を考えよう。すなわち

$K = \{k = (k_1, \dots, k_d); k_i = 0, 1, \dots, p-1\}$, $|K| = p^d$ 個の分集団からなる。全ての $k \in K$ について $N_k = N$ とする。 $k, j \in K$ に対して、それらの和 $k+j$ と、差 $k-j$ を p を法として $(k \pm j)_i = k_i \pm j_i \pmod{p}$ で定義する。移住率は空間的に一様で $k \in K$ から $j \in K$ への移住率は差 $k-j = i$ のみに依存すると仮定する。すなわち $m_{k,j} = m_{k-j} = m_i$ と書くことにする。

このとき 2 個サンプル遺伝子に対する母関数 $\Psi(\alpha)$ を求める。サンプルした分集団を $k, j \in K$ とすると $\alpha = \varepsilon^k + \varepsilon^j$ 、かつ空間的一様性より $\Psi(\alpha) = \Psi(k, j)$ は差 $k-j = i$ にしか依存しない。従って $\Psi(\alpha) = \Psi(k, j) = \Psi(k-j) = \Psi(i)$ と書くことにする。(6.7) より次式が成り立つ。

$$\sum_{j \in K} (m_j + m_{-j}) \Psi(k-j) + \frac{1}{2N} (1 - \Psi(k)) \delta_{k,0} = \lambda \Psi(k)$$

この方程式については Wilkinson-Herbots(1998) により解かれている。

ここでは次の様な $\{\Psi(k); k \in K\}$ の母関数を導入することにより容易に解くことができることを紹介しよう (Notohara(2006))。

d次元ベクトル $\Theta = (\theta_1, \dots, \theta_d)$ を各成分が $\theta_r = \frac{2\pi q}{p}$, ($q = 0, 1, \dots, p-1$) の値を取るものとする。

このように定義されるパラメーター Θ の集合を Ω とする。母関数を

$H(\Theta) = \sum_{k \in K} \Psi(k) \exp[i\Theta \cdot k]$, $i = \sqrt{-1}$ $\Theta \cdot k = \sum_{r=1}^d \theta_r k_r$ で定義する。各 $\Psi(k)$ は逆変換に

より $\Psi(k) = \frac{1}{p^d} \sum_{\Theta \in \Omega} H(\Theta) \exp[-i\Theta \cdot k]$ で与えられる。(6.7) より次式を得る。

$$\{M(\Theta) + M(-\Theta)\} H(\Theta) + \frac{1}{2N} \{1 - \Psi(0)\} = \lambda H(\Theta) \quad (6.8)$$

$$\text{ここで } M(\Theta) = \sum_{k \in K} m_k \exp[i\Theta \cdot k].$$

$$\text{これより } H(\Theta) = \frac{\Psi(0) - 1}{2N\{M(\Theta) + M(-\Theta) - \lambda\}} \quad (6.9)$$

逆変換により

$$\Psi(0) = \frac{1}{p^d} \sum_{\Theta} H(\Theta) = \frac{\Psi(0) - 1}{2Np^d} \left\{ \sum_{\Theta} \frac{1}{M(\Theta) + M(-\Theta) - \lambda} \right\}$$

この式より $\Psi(0)$ を求め、上記の結果を利用すると次の表現式を得る。

$$\Psi(k) = \left\{ \sum_{\Theta} \frac{\exp[-i\Theta \cdot k]}{M(\Theta) + M(-\Theta) - \lambda} \right\} / \left\{ \sum_{\Theta} \frac{1}{M(\Theta) + M(-\Theta) - \lambda} - 2Np^d \right\} \quad (6.10)$$

平均、分散は

$$E(k) = E[T | \alpha = \varepsilon^0 + \varepsilon^k] = 2Np^d + \Gamma_1(k) - \Gamma_1(0)$$

$$V(k) = E[T^2 | \alpha = \varepsilon^0 + \varepsilon^k] - (E(k))^2 = 2\{\Gamma_2(0) - \Gamma_2(k)\} + (E(k))^2 - 2\Gamma_1(k)E(k)$$

$$\text{ここで } \Gamma_1(k) = \sum_{\Theta} \frac{\exp[-i\Theta \cdot k]}{M(\Theta) + M(-\Theta)}, \quad \Gamma_2(k) = \sum_{\Theta} \frac{\exp[-\Theta \cdot k]}{\{M(\Theta) + M(-\Theta)\}^2}$$

簡単な「島モデル」で上の結果を見てみよう。p 個の分集団が 1 次元円周上に配置され、

移住率は等しく $m_k = \begin{cases} m & \text{if } k \neq 0 \\ -(p-1)m & \text{if } k = 0 \end{cases}$ 、また集団サイズは簡単のため $2N = 1$ とす

る。このとき $M(\Theta) = pm(\delta(\Theta) - 1)$ ただし $\Theta = \frac{2\pi q}{p}$, $q = 0, 1, \dots, p-1$ 。

$$\Psi(k) = \begin{cases} \frac{2m + \lambda}{2m + (2pm + 1)\lambda + \lambda^2} & (k = 0) \\ \frac{2m}{2m + (2pm + 1)\lambda + \lambda^2} & (k \neq 0) \end{cases} \quad (6.11)$$

$$\lambda^2 + (2pm + 1)\lambda + 2m = 0 \text{ の解を } \lambda_1 = \frac{-(2pm + 1) + \sqrt{(2pm + 1)^2 - 8m}}{2},$$

$$\lambda_2 = \frac{-(2pm + 1) - \sqrt{(2pm + 1)^2 - 8m}}{2} \text{ とすると、 } \Psi(0) = \frac{A_1}{\lambda - \lambda_1} + \frac{A_2}{\lambda - \lambda_2}, k \neq 0 \text{ のとき}$$

$$\Psi(k) = \frac{B_1}{\lambda - \lambda_1} + \frac{B_2}{\lambda - \lambda_2} \text{。 } k = 0, k \neq 0 \text{ のときの合祖時間の分布密度をそれぞれ}$$

$f_0(t), f_1(t)$ とするとラプラス逆変換により次の結果を得る。

$$f_0(t) = A_1 \exp(\lambda_1 t) + A_2 \exp(\lambda_2 t), \quad f_1(t) = B_1 \exp(\lambda_1 t) + B_2 \exp(\lambda_2 t)$$

ただし $A_1 = \frac{\lambda_1 + 2m}{\lambda_1 - \lambda_2}$, $A_2 = \frac{\lambda_2 + 2m}{\lambda_2 - \lambda_1}$ および $B_1 = -B_2 = \frac{2m}{\lambda_1 - \lambda_2}$ である。

地理的構造を持つ遺伝子系図過程 $\alpha(t)$ について、状態 α での滞在時間を $\tau(\alpha)$ とすると $\tau(\alpha)$ はパラメーター $|R_{\alpha,\alpha}|$ の指数分布に従う ((6.4)式)。この滞在時間の後、状態 α から β ($\beta \neq \alpha$) へのジャンプが確率 $\frac{R_{\alpha,\beta}}{|R_{\alpha,\alpha}|}$ で起こる。このマルコフ連鎖をシミュレーションによって再現し、共通な一つの祖先へ到達するまでの待ち時間、合祖時間の分布を求めたのが下の図である。ここでは5個の分集団がサークル上に配置され隣接する分集団間で等しい移住率 m をもつモデルで、5個のサンプル遺伝子についてシミュレーションを行った (Notohara and Umeda(2006))。図 a は一つの分集団から5個の遺伝子をサンプルした場合、図 b は5個の分集団からそれぞれ1個の遺伝子、計5個をサンプルした場合の合祖時間の分布のシミュレーション結果を示している。

図 a

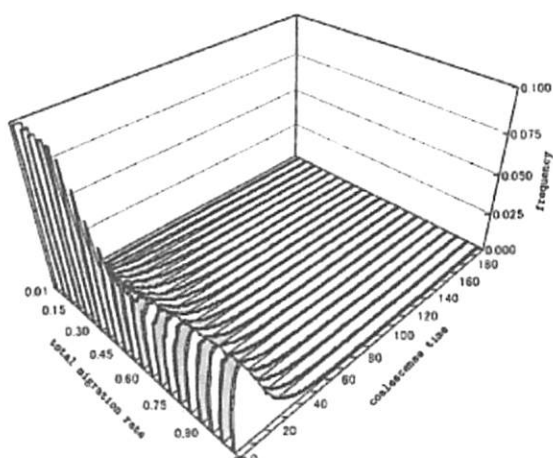
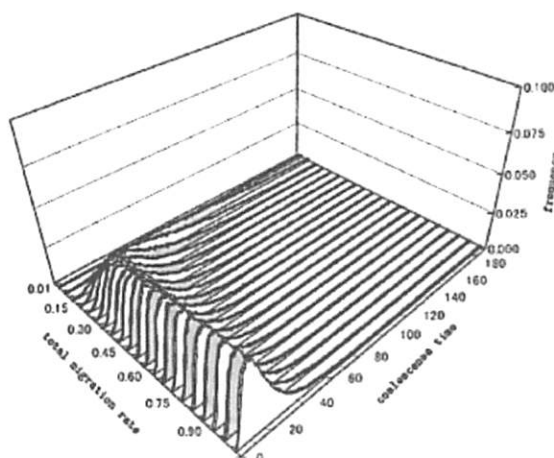


図 b



横軸は合祖時間 T 、縦軸は総移住率 $2m$ を表している。総移住率が 1.0 になると、サンプルの方法によらず、図 a、図 b 同じような分布になることが分かる。他方移住率が小さくなるにつれて、両図の違いが大きくなることが分かる。

第3章で導入した、DNA塩基配列を比較したときこれらの間に見出される分離サイトの数について考察する。 $S(\alpha)$ をサンプル α の中に存在する分離サイトの数とする。無限サイトモデルの下ではこの数は系図上に起こった突然変異の総数に等しい。合祖時間の場合と同様に分離サイトの数 $S(\alpha)$ の母関数 $h(\alpha)$ を $h(z, \alpha) = E_\alpha[z^{S(\alpha)}]$ (ただし $0 \leq z \leq 1$) で定義すると次の定理が成り立つ。

定理 6. 4

$h(z, \alpha) = E[z^{S(\alpha)} | \alpha]$ ($\alpha \in I, 0 \leq z \leq 1$) は次式を満たす (Notohara(1997))。

$$\sum_{\beta \in I} R_{\alpha, \beta} h(z, \beta) = \frac{1}{2} |\alpha| \mathcal{G}(1-z) h(z, \alpha) \quad \text{境界条件} \quad h(z, \varepsilon^k) = 1 \quad (k \in K) \quad (6.12)$$

(証明) 状態 α での滞在時間を $\tau(\alpha)$ とする。 $\tau(\alpha)$ から生成される σ -field を F_τ とすると強マルコフ性 (付録 I 参照) を用いて、

$$h(z, \alpha) = E_\alpha [z^{S(\alpha)}] = E_\alpha [E[z^{S(\alpha)} | F_\tau]] = E_\alpha [E_{\alpha(\tau)} [z^{S(\alpha)}]]。 \text{ サンプル } \alpha \text{ の系図過程で最初に}$$

ジャンプする時刻までに生じる突然変異の数を $X(\alpha)$ とすると $S(\alpha) = S(\alpha(\tau)) + X(\alpha)$ 。

$\tau(\alpha) = t$ のとき $X(\alpha)$ は系図の $|\alpha|$ 本の枝に独立に生じる突然変異の数なので、平均 $\frac{1}{2} \mathcal{G}|\alpha|t$

のポアソン分布に従う。また $\tau(\alpha)$ はパラメーター $|R_{\alpha, \alpha}|$ の指数分布に従う。これより、

$$\begin{aligned} h(z, \alpha) &= E_\alpha [E_{\alpha(\tau)} [z^{S(\alpha(\tau)) + X(\alpha)}]] = E_\alpha [z^{X(\alpha)} E_{\alpha(\tau)} [z^{S(\alpha(\tau))}]] = E_\alpha [z^{X(\alpha)} h(z, \alpha(\tau))] \\ &= \left\{ \sum_{\beta \neq \alpha} \frac{R_{\alpha, \beta}}{|R_{\alpha, \alpha}|} h(z, \beta) \right\} E_\alpha [z^{X(\alpha)}] \end{aligned}$$

$$\begin{aligned} E_\alpha [z^{X(\alpha)}] &= \int_0^\infty \left(\sum_{k=0}^\infty z^k \times \frac{(\mathcal{G}|\alpha|t/2)^k}{k!} \exp(-\frac{1}{2} \mathcal{G}|\alpha|t) \right) \times |R_{\alpha, \alpha}| \exp(-|R_{\alpha, \alpha}|t) dt \\ &= |R_{\alpha, \alpha}| \int_0^\infty \exp\left(-\frac{1}{2} \mathcal{G}|\alpha|(1-z)t\right) \times \exp(-|R_{\alpha, \alpha}|t) dt = \frac{|R_{\alpha, \alpha}|}{(\mathcal{G}|\alpha|(1-z)/2 + |R_{\alpha, \alpha}|)} \end{aligned}$$

$$\text{よって } h(z, \alpha) = \frac{-R_{\alpha, \alpha}}{\frac{1}{2} \mathcal{G}|\alpha|(1-z) - R_{\alpha, \alpha}} \left\{ \sum_{\beta \neq \alpha} \frac{R_{\alpha, \beta}}{-R_{\alpha, \alpha}} h(z, \beta) \right\}, \text{ 整理すると (6.12) を得る。}$$

6.3 大きな移住率および小さな移住率における遺伝子系図モデル

前節で求めた方程式を一般に解くことは困難である。そこで移住が非常に強い場合と弱い場合について考えてみよう。移住の強さを表すパラメーター ζ を用いて、後ろ向き移住率行列を $\zeta M = (\zeta m_{k,j})$ とする。方程式(6.7)は次式の様になる。

$$\zeta \sum_{k, j \in K} \alpha_k m_{k,j} \Psi(\alpha - \varepsilon^k + \varepsilon^j) + \sum_{k \in K} \frac{\alpha_k (\alpha_k - 1)}{4N_k} \{\Psi(\alpha - \varepsilon^k) - \Psi(\alpha)\} = \lambda \Psi(\alpha) \quad (6.13)$$

$$\text{境界条件} \quad \Psi(\varepsilon^k) = 1, \quad k \in K。$$

$\tilde{\alpha}(t)$ を後ろ向き移住率 $M = (m_{k,j})$ によって生成される独立な n 個の粒子のランダムウ

オークを表現するマルコフ連鎖とする。すなわち、 $\tilde{\alpha}(t)$ は次の生成作用素 $\tilde{R} = (\tilde{R}_{\alpha, \beta})$ に

よって生成される。

$$\tilde{R}_{\alpha,\beta} = \begin{cases} \alpha_k m_{k,j} & (\beta = \alpha - \varepsilon^k + \varepsilon^j \text{のとき}) \\ \sum_{k \in K} \alpha_k m_{k,k} & (\beta = \alpha \text{のとき}) \\ 0 & (\text{その他}) \end{cases}$$

移住率 $\zeta M = (\zeta m_{k,j})$ に従って動く独立な n 個の粒子のマルコフ連鎖は $\tilde{\alpha}(\zeta t)$ と表現されることを注意しておこう。

(1) 強移住率極限 (Strong-migration limit)

移住の強さを表すパラメーター ζ が無限に大きくなった極限を強移住率極限と呼ぶことにすると次の定理を得る。

定理 6. 5

$$\lim_{\zeta \rightarrow \infty} \Psi(\alpha) = \prod_{j=2}^n \frac{j(j-1)}{j(j-1) + 4N_e \lambda} \quad \lim_{\zeta \rightarrow \infty} h(\alpha) = \prod_{j=2}^n \frac{j-1}{j-1 + \mathcal{G}^*(1-s)} \quad (6.14)$$

ここで $N_e = \frac{1}{\sum_{k \in S} (\nu_k^2 / N_k)}$ 、 $\mathcal{G}^* = 2N_e \mathcal{G}$ である。 $\nu = (\nu_k; k \in S)$ は $M = (m_{k,j}; k, j \in K)$ から生成される K 上のマルコフ連鎖の定常分布で $\sum_{k \in S} \nu_k = 1$ 、 $\sum_{k \in S} \nu_k m_{k,j} = 0$ を満たす。

(証明)

まず $\Psi(\alpha)$ について証明する。 $\zeta = 1$ としよう。このとき(6.13)は次のように書ける。

$$\sum_{k,j \in K} \alpha_k m_{k,j} \Psi(\alpha - \varepsilon^k + \varepsilon^j) - f(\alpha) \Psi(\alpha) = -h(\alpha)$$

$$\text{ただし、} \quad f(\alpha) = \sum_{k \in K} \frac{\alpha_k (\alpha_k - 1)}{4N_k} + \lambda, \quad h(\alpha) = \sum_{k \in K} \frac{\alpha_k (\alpha_k - 1)}{4N_k} \Psi(\alpha - \varepsilon^k).$$

この方程式の解はファインマン=カツツの公式 (付録(J)参照) を使って次のように表現される。

$$\Psi(\alpha) = \tilde{E} \left[\int_0^\infty h(\tilde{\alpha}(t)) \exp \left(- \int_0^t f(\tilde{\alpha}(s)) ds \right) dt \right]$$

ここで $\tilde{E}[\cdot]$ はマルコフ連鎖 $\tilde{\alpha}(t)$ に関する期待値を意味する。

一般の ζ に対する式(6.13)は $\zeta \sum_{k,j \in K} \alpha_k m_{k,j} \Psi(\alpha - \varepsilon^k + \varepsilon^j) - f(\alpha) \Psi(\alpha) = -h(\alpha)$ と表さ

れ、その解は $\Psi(\alpha) = \tilde{E} \left[\int_0^\infty h(\tilde{\alpha}(\zeta t)) \exp \left(- \int_0^t f(\tilde{\alpha}(\zeta s)) ds \right) dt \right]$ と表現される。

マルコフ連鎖に関する大数の法則より

$$\lim_{\zeta \rightarrow \infty} \frac{1}{t} \int_0^t f(\tilde{\alpha}(\zeta s)) ds = \lim_{\zeta \rightarrow \infty} \frac{1}{\zeta t} \int_0^{\zeta t} f(\tilde{\alpha}(u)) du = \sum_{\beta \in I(n)} P(\beta) f(\beta)$$

$$\text{ここで } I(n) = \{\alpha \in I; |\alpha| = n\}, \quad P(\beta) = \frac{|\beta|!}{\prod_{k \in K} \beta_k!} \prod_{k \in K} \nu_k^{\beta_k}.$$

ただし $\nu = (\nu_k; k \in K)$ は $M = (m_{k,j}; k, j \in K)$ から生成される K 上のマルコフ連鎖の定常

分布であり、 $\sum_{k \in K} \nu_k = 1, \sum_{k \in K} \nu_k m_{k,j} = 0$ を満たす。従って $\{P(\beta); \beta \in I(n)\}$ はマルコフ連鎖

$\{\tilde{\alpha}(t); |\alpha(0)| = n\}$ の定常分布である。 $f(\alpha) = \sum_{k \in K} \frac{\alpha_k(\alpha_k - 1)}{4N_k} + \lambda$ より

$$\begin{aligned} \sum_{\beta \in I(n)} P(\beta) f(\beta) &= \sum_{\beta \in I(n)} \frac{n!}{\prod_{j \in K} \beta_j!} \prod_{i \in K} \nu_i^{\beta_i} \left\{ \sum_{k \in K} \frac{\beta_k(\beta_k - 1)}{4N_k} + \lambda \right\} \\ &= n(n-1) \sum_{k \in K} \left\{ \frac{\nu_k^2}{4N_k} \left(\sum_{\beta_k \geq 2} \frac{(n-2)!}{\left(\prod_{j \neq k} \beta_j! \right) (\beta_k - 2)!} \left(\prod_{i \in K} \nu_i^{\beta_i} \right) \nu_k^{\beta_k - 2} \right) + \lambda \right\} \\ &= n(n-1) \sum_{k \in K} \frac{\nu_k^2}{4N_k} + \lambda = \frac{n(n-1)}{4N_e} + \lambda \end{aligned}$$

$\sum_{\beta_k \geq 2}$ は条件 $\beta_k \geq 2$ を満たす全ての $\beta \in I(n)$ についての和を意味する。以上の準備の上で、

まず $|\alpha| = 2$ のとき、 $\lim_{\zeta \rightarrow \infty} \Psi(\alpha)$ が初期条件 α に依存しないことを示す。 $\alpha = \varepsilon^k + \varepsilon^j$ とする。

全ての $k \in K$ について $\Psi(\varepsilon^k) = 1$ より、

$$h(\alpha) = \sum_{i \in K} \frac{1}{2N_i} \delta(\alpha, 2\varepsilon^i) = \begin{cases} \frac{1}{2N_k} & (k = j \text{ のとき}) \\ 0 & (\text{その他}) \end{cases}$$

$$\begin{aligned} \lim_{\zeta \rightarrow \infty} \Psi(\alpha) &= \tilde{E} \left[\int_0^\infty h(\tilde{\alpha}(\infty)) \exp \left(- \left(\frac{1}{2N_e} + \lambda \right) t \right) dt \right] \\ &= \int_0^\infty \tilde{E} \left[\sum_{i \in K} \frac{1}{2N_i} \delta(\tilde{\alpha}(\infty), 2\varepsilon^i) \right] \exp \left(- \left(\frac{1}{2N_e} + \lambda \right) t \right) dt \end{aligned}$$

$\tilde{\alpha}(\infty)$ は定常分布 $\{P(\beta); \beta \in I(n)\}$ に従う確率変数であり、これより

$$\tilde{E}\left[\sum_{i \in K} \frac{1}{2N_i} \delta(\tilde{\alpha}(\infty), 2\varepsilon^i)\right] = \sum_{i \in K} \frac{\nu_i^2}{2N_i}. \text{ 故に}$$

$$\lim_{\zeta \rightarrow \infty} \Psi(\alpha) = \int_0^\infty \left(\sum_{i \in K} \frac{\nu_i^2}{2N_i} \right) \exp\left(-\left(\frac{1}{2N_e} + \lambda\right)t\right) dt = \frac{1/2N_e}{(1/2N_e) + \lambda}$$

すなわち、極限 $\lim_{\zeta \rightarrow \infty} \Psi(\alpha)$ ($\alpha \in I(2)$) は α に依存しない。そこでこの極限を $|\alpha| = 2$ より

単に $\lim_{\zeta \rightarrow \infty} \Psi(\alpha) = \Psi(2)$ で表すことにする。 $|\alpha| \geq 3$ の場合についてはファイマン=カツツ

の公式 $\Psi(\alpha) = \tilde{E}\left[\int_0^\infty h(\tilde{\alpha}(t)) \exp\left(-\int_0^t f(\tilde{\alpha}(s)) ds\right) dt\right]$ において

$$f(\alpha) = \sum_{k \in K} \frac{\alpha_k(\alpha_k - 1)}{4N_k} + \lambda, \quad h(\alpha) = \sum_{k \in K} \frac{\alpha_k(\alpha_k - 1)}{4N_k} \Psi(\alpha - \varepsilon^k) \text{ より上式が } \Psi(\alpha) \text{ に関する}$$

漸化式になっていることが分かる。よって数学的帰納法によって証明が完了する。

分離サイトの数 $S(\alpha)$ の母関数 $h(z, \alpha) = E_\alpha[z^{S(\alpha)}]$ の強移住率極限 (Strong-migration limit) についても、同様の議論により証明される (Notohara(1997))。 (証明終わり)

N_e は強移住率極限の下での有効個体数であるが、定理 6. 5 での N_e の定義式は加重 $\{\nu_k; k \in K\}$ を付けた $\left\{\frac{N_k}{\nu_k}; k \in K\right\}$ の調和平均と見ることができる。調和平均より相加平均

$$\text{が大であることより } N_e = \frac{1}{\sum_{k \in K} (\nu_k^2 / N_k)} \leq \sum_{k \in K} \nu_k \left(\frac{N_k}{\nu_k}\right) = \sum_{k \in K} N_k = N_T \text{ が成り立つ。}$$

地理的構造を持つ場合は移住率が無限に大きくなった任意交配集団であっても、有効個体数は全集団の実個体数に比べて小さいことを意味する。

移住率がかなり大きいとき、強移住率極限で用いた大数の法則の代わりに、中心極限定理を利用することにより、移住率が大きいときの Coalescence time の近似分布を求めることができる。保存的移住率の場合、全集団は有効個体数が強移住率極限のときの有効個体数 N_e を移住率で修正した N_e^* を有効個体数とするほとんど任意交配集団に近い集団と見なすことができる (Notohara(2000, 2010))。簡単な例を紹介しよう。

<例> d次元トールモデル:

$$K = \{k = (k_1, \dots, k_d); k_i = 1, 2, \dots, p\}, |K| = p^d \quad N_e = Np^d$$

もし移住の強度 ζ が非常に大きなとき、近似的に次式が成り立つ。

$$\Psi(\alpha) \approx \prod_{j=2}^n \frac{j(j-1)}{j(j-1) + 4N_e^* \lambda}$$

$$\text{ここで } N_e^* = N_e - \frac{M_0^*}{2\zeta}, M_0^* = \sum_{\theta \neq 0} \frac{1}{M(\theta) + M(-\theta)}, \text{かつ } M(\theta) = \sum_{k \in S} m_k \exp(ik \cdot \theta).$$

$$\text{平均、分散は } E[T|\alpha] \approx 4N_e^* \sum_{k=2}^n \frac{1}{k(k-1)}, \quad \text{Var}[T|\alpha] \approx (4N_e^*)^2 \sum_{k=2}^n \frac{1}{k^2(k-1)^2} \text{となる。}$$

この中で特に簡単な2つのモデルを紹介すると

① p 個の分集団から成る島モデル

$$\text{移住率を } m_{k,j} = \begin{cases} 1 & \text{if } k \neq j \\ -(p-1) & \text{if } k = j \end{cases}, \text{この時 } N_e^* = Np + \frac{p-1}{4p\zeta}$$

$$\text{全移住率} = (p-1)\zeta$$

② p 個の分集団から成る環状飛び石モデル

$$\text{移住率を } m_{k,j} = \begin{cases} 1 & \text{if } k = j \pm 1 \pmod{p} \\ -2 & \text{if } k = j \end{cases}, \text{この時 } N_e^* = Np + \frac{p^2-1}{48\zeta}$$

$$\text{全移住率} = 2\zeta$$

非保存的移住率の場合、Coalescence time の分布は保存的移住の場合と異なり、移住率が大きい場合でも、近似分布が任意交配とは異なる分布になることが示される。詳しくは Notohara(2010)を参照されたい。

(2) 弱移住率極限 (Weak-migration limit)

移住率が逆に非常に弱い場合を考える。この場合、異なる分集団から取り出した遺伝子の合祖時間は移住率が小さくなるにつれ増加する。従って、 $\zeta \rightarrow 0$ とすると合祖時間分布の通常の極限は存在しない。そこで移住の強度でスケーリングした ζT の分布を考える。 ζT のラプラス変換を $\Phi(\alpha) = E[\exp(-\lambda(\zeta T))|\alpha]$ とすると次式が成り立つ。

$$\zeta \sum_{k,j \in K} \alpha_k m_{k,j} \Phi(\alpha - \varepsilon^k + \varepsilon^j) + \sum_{k \in K} \frac{\alpha_k(\alpha_k - 1)}{4N_k} \{\Phi(\alpha - \varepsilon^k) - \Phi(\alpha)\} = \lambda \zeta \Psi(\alpha) \quad (6.15)$$

$$\text{境界条件: 全ての } k \in K \text{ について } \Phi(\varepsilon^k) = 1.$$

定理 6. 6

$\Phi(\alpha)$ の弱移住極限を $\Phi^*(\alpha) = \lim_{\zeta \rightarrow 0} \Phi(\alpha)$ ($\alpha \in I$) とすると次の様な関係を満たす。

(i) $\alpha = n\varepsilon^k$ ($n \geq 1$) (ただし $k \in K$) ならば $\Phi^*(\alpha) = 1$.

すなわち、極限分布は原点に確率 1 をもつデルタ分布となる。

(ii) $\alpha = n_1\varepsilon^{k^1} + n_2\varepsilon^{k^2} + \dots + n_r\varepsilon^{k^r}$ 、ただし $n_i \geq 1$ ($1 \leq i \leq r$) かつ $i \neq j$ のとき

$$k^i \neq k^j \text{ ならば } \Phi^*(\alpha) = \Phi^*(\varepsilon^{k^1} + \dots + \varepsilon^{k^r}).$$

(iii) $I_1 = \{\alpha = (\alpha_k; k \in K) \in I; \alpha_k \leq 1 \text{ for all } k \in K\}$ とする。 $\alpha \in I_1$ ならば

$$\text{次の式を満たす。 } \sum_{k, j \in K} \alpha_k m_{k, j} \Phi^*(\alpha - \varepsilon^k + \varepsilon^j) = \lambda \Phi^*(\alpha) \quad (\text{Notohara(2001)})$$

これら三つの関係式を用いて $\{\Phi^*(\alpha); \alpha \in I\}$ を解くことができる。

<例>

① p 分集団からなる島モデル

$$\Phi^*(\alpha) = \prod_{k=2}^r \frac{k(k-1)}{k(k-1) + \lambda} \quad \text{ただし } \alpha = n_1\varepsilon^{k^1} + \dots + n_r\varepsilon^{k^r}, r \geq 2 \quad (6.16)$$

Kingman の Coalescent モデルの場合と同型の母関数となる。

これより、移住率強度 ζ が非常に小さいとき、 $\Psi(\alpha)$ の近似解として

$$\Psi(\alpha) \approx \prod_{k=2}^r \frac{k(k-1)\zeta}{k(k-1)\zeta + \lambda} \quad \text{となる。}$$

② 5 個の分集団から成る環状飛び石モデル $K = \{1, 2, 3, 4, 5\}$ とする。このとき、

$$\Phi^*(1, 2) = \frac{2(\lambda + 2)}{\lambda^2 + 6\lambda + 4}, \quad \Phi^*(1, 2, 3) = \frac{4(\lambda^2 + 8\lambda + 20)}{(\lambda^2 + 10\lambda + 20)(\lambda^2 + 6\lambda + 4)} \quad (6.17)$$

$$\Phi^*(1, 2, 3, 4, 5) = \frac{80}{(\lambda^2 + 10\lambda + 20)(\lambda^2 + 6\lambda + 4)} \quad \text{となる。}$$

下図は図 a, b と同じサークル状モデルで、 ζ によってスケーリングされた合祖時間 ζT のシミュレーション結果を示している。図 c は一つの分集団から 5 個の遺伝子をサンプルした場合、図 d は各集団から 1 個ずつ遺伝子をサンプルした場合の結果である (Notohara & Umeda (2006))。 $\zeta \rightarrow 0$ とすると、図 c ではデルタ分布に収束し、図 d では弱移住率極限の分布 ((6.16) 式、 $p=r=5$) に収束していることが分かる。

図 c

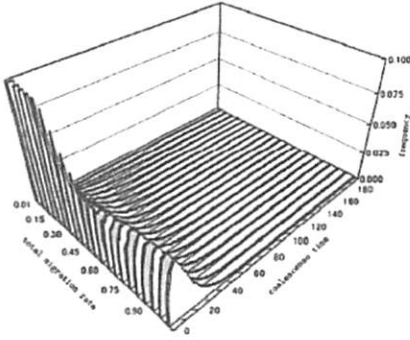
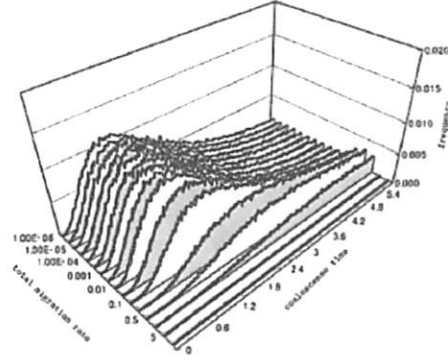


図 d



6. 4 分離サイト数からの合祖時間の推定

任意交配集団における DNA 塩基配列中の分離サイトの数と合祖時間の関係を

3. 2 節で述べたが、この方法を地理的構造を持つ集団に拡張してみよう。サンプル α に対して合祖時間 T と分離サイト S の同時母関数を

$$G(\alpha; \lambda, z) = E[\exp(-\lambda T) z^S | \alpha] : (\lambda \geq 0, 0 \leq z \leq 1) \text{ で定義する。}$$

第 3. 2 節と同じく、無限サイトモデルを仮定し、突然変異は全系図長 $L = \sum_{j=2}^n j \tau_j$

($n = |\alpha|$.) の系図上に平均 $\frac{\theta L}{2}$ のポアソン分布に従って起こるものとする。これより

次の表現式も得られる。

$$\begin{aligned} G(\alpha; \lambda, z) &= E[\exp(-\lambda T) E[z^S | \tau_2, \dots, \tau_n]] = E[\exp(-\lambda T) \sum_{k=0}^{\infty} \left\{ \frac{(\frac{\theta L}{2})^k}{k!} \exp(-\frac{\theta L}{2}) \right\} z^k] \\ &= E[\exp\{-\sum_{j=2}^n (\frac{\theta(1-z)j}{2} + \lambda) \tau_j\} | \alpha] \end{aligned} \quad (6.18)$$

$\tau(\alpha)$ を状態 $\alpha_0 = \alpha$ からの最初のジャンプ時刻とすると、マルコフ時間 $\tau(\alpha)$ は指数分布

$P(\tau(\alpha) \geq t) = \exp(-|R_{\alpha, \alpha}| t)$ に従うので、強マルコフ性 (付録 (I) 参照) を使って

$$G(\alpha; \lambda, z) = \sum_{\beta \neq \alpha} \frac{R_{\alpha, \beta}}{|R_{\alpha, \alpha}|} \times \frac{|R_{\alpha, \alpha}|}{\frac{\theta}{2}(1-z)|\alpha| + \lambda + |R_{\alpha, \alpha}|} G(\beta; \lambda, z). \quad \text{これより}$$

$$\sum_{\beta \in I} R_{\alpha, \beta} G(\beta; \lambda, z) = \left\{ \lambda + \frac{\theta}{2}(1-z)|\alpha| \right\} G(\alpha; \lambda, z). \quad (6.19)$$

移住率行列を $\zeta M = (\zeta m_{k,j})$ として強移住率極限 ($\zeta \rightarrow \infty$) および弱移住率極限 ($\zeta \rightarrow 0$)

を考えよう。

(1) 強移住率極限

$$\lim_{\zeta \rightarrow \infty} G(\alpha; \lambda, z) = \prod_{j=2}^n \left(\frac{j(j-1)}{2N_e g(1-z)j + 4N_e \lambda + j(j-1)} \right), \text{ ただし } N_e = \left(\sum_{k \in S} \frac{v_k^2}{N_k} \right)^{-1}.$$

この結果は定理 3. 5 の任意交配集団における母関数と一致する。

(2) 弱移住率極限.

ζ でスケーリングした $(T^*, S^*) = (\zeta T, \zeta S)$ の母関数を次式で定義する。

$$F(\alpha; \lambda, z) = E[\exp(-\lambda \zeta T) z^{\zeta S} | \alpha] \quad (6.20)$$

強マルコフ性により次式が成り立つ。

$$\sum_{\beta \in I} R_{\alpha, \beta} F(\beta; \lambda, z) = \left\{ \zeta \lambda + \frac{g}{2} (1-z^\zeta) | \alpha \right\} F(\alpha; \lambda, z) \quad (6.21)$$

弱移住率極限を $F^*(\alpha) = \lim_{\zeta \rightarrow 0} F(\alpha; \lambda, z)$ とすると,

(i) $\alpha = n\varepsilon^k$ ($n \geq 1$) (ただし $k \in K$) ならば $F^*(\alpha) = 1$.

(ii) $\alpha = n_1 \varepsilon^{k^1} + n_2 \varepsilon^{k^2} + \dots + n_r \varepsilon^{k^r}$ 、ただし $n_i \geq 1$ ($1 \leq i \leq r$) かつ $i \neq j$ のとき

$k^i \neq k^j$ ならば $F^*(\alpha) = F^*(\varepsilon^{k^1} + \dots + \varepsilon^{k^r})$ 。

(iii) $\alpha \in I_1$, ならば $\sum_{k, j \in K} \alpha_k m_{k, j} F^*(\alpha - \varepsilon^k + \varepsilon^j) = \left(\lambda - \frac{1}{2} g | \alpha | \log z \right) F^*(\alpha)$.

これらの方程式より順次弱移住率極限 $F^*(\alpha)$ を求めることができる。

例: p 個の分集団からなる島モデル

$$F^*(\alpha) = \prod_{k=2}^r \left(\frac{k(k-1)}{k(k-1) + \lambda - \frac{1}{2} k g \log z} \right) \quad (\text{ただし } \alpha = n_1 \varepsilon^{k^1} + \dots + n_r \varepsilon^{k^r})$$

ここで $-\log z = \eta$ と置くと

$$F^*(\alpha) = \prod_{k=2}^r \left(\frac{k(k-1)}{k(k-1) + \lambda + \frac{1}{2} k g \xi} \right) = \lim_{\zeta \rightarrow 0} E[\exp(-\zeta \lambda T - \zeta \xi S) | \alpha].$$

$r=2$ のとき, $F^*(\alpha) = \frac{2}{2 + \lambda + g \xi}$. (T^*, S^*) の同時分布密度 (Pd) は

$Pd(T^* = t, S^* = s) = 2 \exp(-2t) \delta(g t - s)$ ただし $\delta(\cdot)$ は Dirac のデルタ関数.

$Pd(S^* = s) = \frac{2}{g} \exp(-\frac{2}{g} s)$ T^* の条件付分布は

$$Pd(T^* = t | S^* = s) = \frac{Pd(T^* = t, S^* = s)}{Pd(S^* = s)} = \delta(t - \frac{s}{g}) \text{ となる。}$$

$r \geq 3$, のとき

$$Pd(T^* = t, S^* = s) = \frac{2}{r g} \sum_{k=2}^r \sum_{j=2}^r (-1)^{k-j} (k-j)_r C_k \times_r C_j \times k(k-1)j(j-1) \times \exp[-k(k-1)t - \frac{2}{g}(k+j-1)\chi(s - \frac{1}{2}kgt)]$$

$$\text{ただし } \chi(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x \geq 0 \end{cases}$$

$$Pd(S^* = s) = \frac{2(r-1)}{g} \sum_{k=2}^r (-1)^{k-2} {}_{r-2}C_{k-2} \exp[-\frac{2(k-1)}{g}s] \text{ より } T^* \text{ の条件付分布は}$$

$$Pd(T^* = t | S^* = k) = \frac{Pd(T^* = t, S^* = k)}{Pd(S^* = k)}.$$

最後に(6.19)に戻り、簡単な具体例で、サンプル数 2 の場合の分離サイト数 S を与えたとき合祖時間 T を求める条件付き確率を求めてみよう。

<例> d 次元トーラス状格子空間のモデル —サンプル数 2 の場合—

島モデルとサークル状の飛び石モデルでサンプル数が 2 の場合の結果を紹介する。

(1) p 個の分集団から成る島モデル(Island model)

$$K = \{1, 2, \dots, p\}, \theta = \frac{2\pi}{p} q \quad (q = 0, 1, \dots, p-1) \text{ および } M(\theta) = pm(\delta(\theta) - 1).$$

$$G(k; \lambda, z) = E[\exp(-\lambda T) z^S | \alpha = \varepsilon^0 + \varepsilon^k] = \frac{x\delta(k) + 2m}{x^2 + (2pm+1)x + 2m}$$

$$\text{ただし } x = g(1-z) + \lambda.$$

$$G(0) = G(0; \lambda, z), \quad G(1) = G(1; \lambda, z)$$

$$\begin{cases} G(0) = \sum_{k=0}^{\infty} \left\{ \frac{A}{(1+\alpha)^{k+1}} \left(\frac{g(1+\alpha)}{g(1+\alpha)+\lambda} \right)^{k+1} + \frac{B}{(1+\beta)^{k+1}} \left(\frac{g(1+\beta)}{g(1+\beta)+\lambda} \right)^{k+1} \right\} z^k \\ G(1) = \sum_{k=0}^{\infty} C \left\{ -\frac{1}{(1+\alpha)^{k+1}} \left(\frac{g(1+\alpha)}{g(1+\alpha)+\lambda} \right)^{k+1} + \frac{1}{(1+\beta)^{k+1}} \left(\frac{g(1+\beta)}{g(1+\beta)+\lambda} \right)^{k+1} \right\} z^k \end{cases}$$

$$\text{ただし } A = \frac{-4m + 2pm + 1 + \sqrt{(2pm+1)^2 - 8m}}{2g\sqrt{(2pm+1)^2 - 8m}}$$

$$B = \frac{4m - 2pm - 1 + \sqrt{(2pm+1)^2 - 8m}}{2g\sqrt{(2pm+1)^2 - 8m}}, \quad C = \frac{2m}{2g\sqrt{(2pm+1)^2 - 8m}} \text{ そして}$$

$$\alpha = \frac{2pm+1+\sqrt{(2pm+1)^2-8m}}{2g}, \quad \beta = \frac{2pm+1-\sqrt{(2pm+1)^2-8m}}{2g}.$$

これより (T, S) の同時分布密度を $Pd(T=t, S=k)$ とすると

$$Pd(T=t, S=k) = \frac{A_1}{(1+\alpha)^{k+1}} \text{Gamma}(k+1, \frac{1}{g(1+\alpha)}) + \frac{A_2}{(1+\beta)^{k+1}} \text{Gamma}(k+1, \frac{1}{g(1+\beta)})$$

$$P(S=k) = \frac{A_1}{(1+\alpha)^{k+1}} + \frac{A_2}{(1+\beta)^{k+1}} \quad \text{および} \quad Pd(T=t|S=k) = \frac{P(T=t, S=k)}{P(S=k)} \quad (6.22)$$

ただし サンプル $\alpha = 2\varepsilon^k$ のとき $(A_1, A_2) = (A, B)$ 、

サンプル $\alpha = \varepsilon^k + \varepsilon^j$ のとき $(A_1, A_2) = (-C, C)$ とする。

(2) サークル状飛び石モデル $K = \{1, 2, \dots, p\}$, $G(k; \lambda, z) = E[\exp(-\lambda T) z^S | \varepsilon^0 + \varepsilon^k]$

$$G(k; \lambda, z) = \left(\sum_{\theta} \frac{\exp(-i\theta k)}{M(\theta) + M(-\theta) - x} \right) / \left(\sum_{\theta} \frac{1}{M(\theta) + M(-\theta) - x} - 2Np \right)$$

ここで $x = g(1-z) + \lambda$, $\theta = \frac{2\pi q}{p}$ ($q = 1, 2, \dots, p$) および $M(\theta) = 2m(\cos \theta - 1)$.

$$G(k; \lambda, z) = \left(\sum_{q=1}^p \frac{\exp(-i \frac{2\pi q}{p} k)}{4m(\cos \frac{2\pi q}{p} - 1) - x} \right) / \left(\sum_{q=1}^p \frac{1}{4m(\cos \frac{2\pi q}{p} - 1) - x} - 2Np \right)$$

ここで分集団の数 p を無限に大きくすると、留数定理により

$$\int_0^{2\pi} \frac{\exp(-ik\theta)}{4m(\cos \theta - 1) - x} d\theta = \frac{-2\pi}{\sqrt{8mx + x^2}} \left(\frac{4mx - \sqrt{8mx + x^2}}{4m} \right)^k \quad \text{なので}$$

$$\begin{aligned} \lim_{p \rightarrow \infty} G(k; \lambda, z) &= \left\{ \int_0^{\infty} \frac{\exp(-ik\theta)}{4m(\cos \theta - 1) - x} d\theta \right\} / \left\{ \int_0^{2\pi} \frac{1}{4m(\cos \theta - 1) - x} d\theta - 4N\pi \right\} \\ &= \left(\frac{1}{1 + 2N\sqrt{8mx + x^2}} \right) \left(\frac{4m + x - \sqrt{8mx + x^2}}{4m} \right)^k, \quad (\text{ただし } x = g(1-z) + \lambda) \end{aligned}$$

を得る(Notohara and Umeda(2006)).

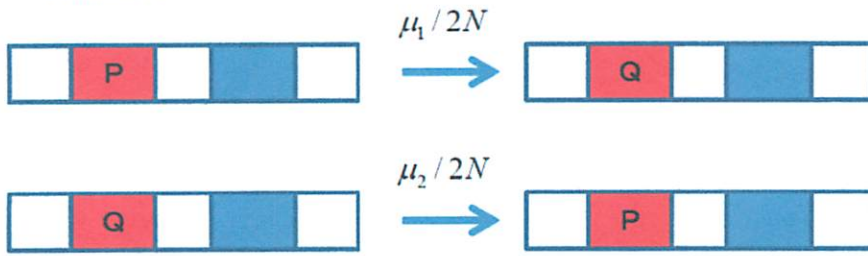
6. 5 自然選択による構造を持つ遺伝子系図 (Hudson&Kaplan のモデル)

構造を持つ遺伝子系図が現れる例としては Kaplan, Darden and Hudson(1988), Hudson and Kaplan(1988)によって導入され、その後 Barton, Ethridge and Strum(2004)によって厳密に証明された対立遺伝子のタイプによる自然選択を考慮に入れたモデルがある。すなわち 2 種の対立遺伝子 P,Q を持つ自然選択を受けている遺伝子座 S とその遺伝子座に連鎖している中立な遺伝子座 L を考える。遺伝子 L の系図あるいは遺伝的多様性を考えるとき、自然選択を受けている遺伝子座 S によってその系図は大きく影響される。任意交配を行っている 2 倍体の雌雄同株の生物集団を仮定しよう。サンプルした遺伝子の系図を考えると、まず突然変異あるいは遺伝子座 S,L 間の組み換えを無視したとき、対立遺伝子 P を持つ遺伝子 (染色体) と対立遺伝子 Q を持つ遺伝子 (染色体) が共通な祖先をもつことはありえない。すなわち遺伝子 L が連鎖した遺伝子座 S の対立遺伝子によって集団が二つ (P 集団と Q 集団) に分割され、各分集団内で系図を遡ることができる。さらに P,Q 間の突然変異および遺伝子座 S,L 間の組み換えを考慮すると中立な遺伝子座 L の連鎖した遺伝子座 S のタイプの乗り換えが発生することが分かる。すなわち中立な遺伝子は、あるときは P と連鎖することにより分集団 P に属し、またあるときは Q と連鎖することにより分集団 Q に属することになる。ただし、分集団の大きさは対立遺伝子 P,Q の頻度に依存するので、集団サイズは確率的に変動することになる。ただし、遺伝子系図は時間を過去に遡って取るので遺伝子座 S の遺伝子頻度の変化も時間を過去に遡って逆進過程で考える必要がある。ライト・フィッシャーモデルを用いる場合、推移確率は二項分布で与えられるが、自然選択、突然変異及び組み換えの存在下で、その定常分布は具体的形を求めることは困難であり、時間を遡る逆進過程を具体的に求めることができない。Hudson and Kaplan(1988)では離散時間モデルから Coalescent モデルへの収束を厳密には導いていないが、Barton、Ethridge and Strum(2004)ではライト・フィッシャーモデルの代わりに Moran モデルを用いることにより、Coalescent 過程への弱収束の証明を与えている。遺伝子頻度変化と構造を持つ合祖過程を含む確率過程となるが、例えば強い自然選択と突然変異により遺伝子頻度の変化が平衡状態としてほとんど時間変化がないとき、近似的に一定の大きさの P 集団と Q 集団からなる 2 分集団の遺伝子系図モデルと等価なモデルになる。詳細は上記の論文を参照していただくことにして、そのモデルのみを紹介する。下図は染色体上の連鎖した二つの S,L 遺伝子座を表している。

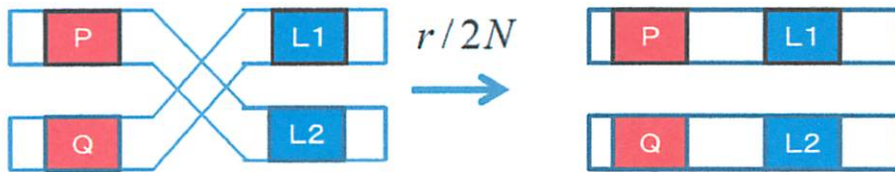


N 個の雌雄同株の 2 倍体生物からなる集団で、遺伝子座 S の対立遺伝子 P,Q 間の突然変異率、 $P \rightarrow Q$ を $\mu_1/2N$ 、 $Q \rightarrow P$ を $\mu_2/2N$ 、遺伝子座 S,L 間の組み換え率を $r/2N$ とする。

突然変異



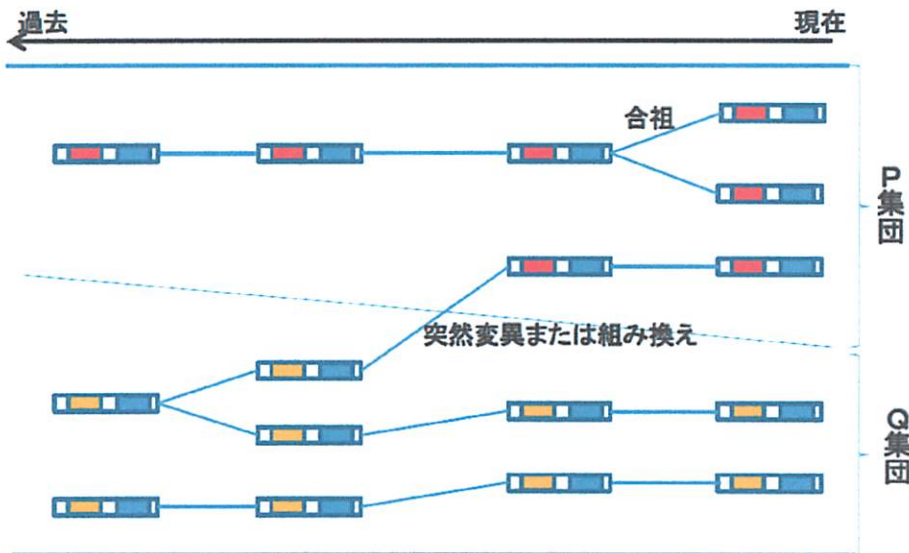
組み換え



自然選択は遺伝子型 PP , PQ , QQ の適応度を $w_{11} = 1 + \frac{s_{11}}{N}$, $w_{12} = 1 + \frac{s_{12}}{N}$,

$w_{22} = 1 + \frac{s_{22}}{N}$ とする。Barton, Ethridge and Strum(2004)では等価な自然選択を含むモ

ランモデルを用いている。この遺伝子系図を表現するためには、遺伝子頻度（対立遺伝子 P の頻度）、 P および Q に連鎖した遺伝子座 L のサンプルおよび祖先遺伝子数の3変数が必要なので状態空間 $E = [0, 1] \times N_0 \times N_0$, $(p, n_1, n_2) \in E$ となる。 n_1, n_2 はそれぞれ P および Q に連鎖した遺伝子座 L のサンプルおよび祖先遺伝子数である。



$N \rightarrow \infty$ の極限において、次の生成作用素 A を持つ E 上の連続時間マルコフ連鎖が得られる。 E 上で定義され実数値、 p に関して 2 階連続微分可能な関数を $f(p, n_1, n_2)$ 、また $q = 1 - p$ とすると

$$\begin{aligned}
 Af(p, n_1, n_2) = & \frac{1}{2p} \binom{n_1}{2} (f(p, n_1 - 1, n_2) - f(p, n_1, n_2)) + \frac{1}{2q} \binom{n_2}{2} (f(p, n_1, n_2 - 1) - f(p, n_1, n_2)) \\
 & + \left(\frac{q\mu_2}{2p} n_1 + \frac{qr}{2} n_1 \right) (f(p, n_1 - 1, n_2 + 1) - f(p, n_1, n_2)) \\
 & + \left(\frac{p\mu_1}{2q} n_2 + \frac{pr}{2} n_2 \right) (f(p, n_1 + 1, n_2 - 1) - f(p, n_1, n_2)) \\
 & + \frac{1}{2} (-\mu_1 p + \mu_2 q + spq) \frac{\partial}{\partial p} f(p, n_1, n_2) + \frac{1}{4} pq \frac{\partial^2}{\partial p^2} f(p, n_1, n_2)
 \end{aligned}$$

Barton, Ethridge and Strum(2004)

Coalescent 過程への収束を示すためには、遺伝子頻度の境界 $p = 0, 1$ の近傍でのシステムの振る舞いを詳細に調べる必要がある。証明の詳細は Barton, Ethridge and Strum(2004)を参照されたい。この論文は自然選択圧などを含む場合の遺伝子系図は分集団サイズが確率的に変動する Structured coalescent model として表現できることを示したとも言える。強い自然選択圧と突然変異による平衡により遺伝子頻度がほとんど一定のとき、近似的に生成作用素 A は通常の Structured Coalescent 過程と同形のモデルになる。Hudson and Kaplan 及び Barton et al のモデルはその後多くの研究(Kaplan et al(1991), Hey(1991), Nordborg(1997, 1999), Nordborg and Krone(2002)等)があるが、特に Selective Sweep (Kaplan et al(1989), Braverman et al(1995), Przeworski(2002)等)あるいは超優性 (Hudson and Kaplan(1988), Begun et al(1999)等) などの研究にも応用されている。