

遺傳的変異の集団遺傳学的解析

九州大学・大学院理学研究院 館田 英典 (Hidenori Tachida)
Department of Biology, Faculty of Sciences
Kyushu University

1 はじめに

生物の進化は集団の遺傳的構成の変化過程とらえることができる。この過程において遺傳子は親から子に伝わるが、次世代が構成される際、進化的要因と呼ばれる突然変異、自然淘汰、集団サイズや移住などを含む集団の構造、が影響を及ぼす。我々が現在観測できるのはこのような過程が長時間積み重なった結果としての、種内・種間の遺傳的変異ということになる。集団遺傳学では様々な進化的要因を仮定してモデルを構築し、種内や種間の変異がどのようになるかを予測し、実際に観測された種内・種間の遺傳的変異と比べることによって、過去に働いた進化的要因について推測をし、生物の進化過程を理解しようとする。

この論文では、まず遺傳子進化のモデルで最も単純な（淘汰に対する）中立・任意交配・一定サイズモデルを仮定し、最近大量にデータが得られるようになったDNA塩基配列の種内・種間変異量とパターンに関して既に知られている結果を簡単に説明する。次にこれらの仮定のうち集団構造に関する仮定が成り立たない場合（集団サイズが増加する場合、集団が別れていて分断・融合を繰り返す場合）について最近我々が得た結果について解説する。

2 中立・任意交配・一定サイズモデル

まず集団遺傳学の standard model である Wright-Fisher モデルについて説明する。世代が不連続で N 個体からなる半数体生物集団のある特定の遺傳子座に着目する。この遺傳子座の遺傳子構成の動態を考える場合、遺傳子座間の相互作用があるような複雑な場合を除くと、 N 個の遺傳子の集団のみに着目し、これらの遺傳子がどのように次世代の遺傳子集団を作るかを考慮すれば良い。中立・任意交配 Wright-Fisher モデルでは、次世代は前の世代の N 個の遺傳子から重複を許して等確率で N 個の遺傳子をサンプルすることによって構成されると仮定する。サンプリングが等確率で行われず、遺傳子のタイプによって異なる場合は自然淘汰が働いている場合であり、また N 個の遺傳子が複

数のグループに分かれており、次世代に同じグループの親から遺伝子が由来する確率と、別のグループから由来する確率に違いがある場合が、集団の構造がある場合である。また次世代が構成される時に遺伝子が増える場合があり、突然変異と呼ばれる。ここでは一遺伝子あたり、一世代あたり突然変異率は u であるとする。

中立・任意交配 Wright-Fisher モデルを仮定し、集団サイズは時間によらず一定であるとした時に、遺伝子系図学理論 (Kingman 1982 または Tavaré 1984 の総説を参照) を使って集団内の変異についていくつかの予想をすることができる。まずある世代 (この世代を $t=0$ とする。また時間はこの時点を起点に過去にさかのぼって計測する) にこの集団から二つの遺伝子をサンプルした場合、この二つの遺伝子が最初の共通祖先を持つ (coalescence と呼ぶ) 時間を T で表し、その分布を求めよう。この二つの遺伝子が直前の世代 ($t=1$) に共通祖先を持つ確率 $\text{Prob}[T=1]$ は、Wright-Fisher モデルの仮定から $1/N$ となり、この余事象の確率 $\text{Prob}[T>1]$ は $1-1/N$ となる。各世代での遺伝子のサンプリングは独立なので、この過程を続けていくと一般に

$$\text{Prob}[T > t] = \left(1 - \frac{1}{N}\right)^t \approx \exp\left[-\frac{t}{N}\right]$$

となるのがわかる。なお最後の近似は $N \gg 1$ の条件のもとで成立するが、実際の生物集団はかなり大きくこの条件は通常満たされているので、今後この近似を使って説明を行う。このことから二つの遺伝子が共通祖先を持つ時間 T は指数分布をすることがわかる。次に n 個の遺伝子をサンプルした時を考える。 n 個の遺伝子のどれもが前の世代に共通祖先を持たない確率は上と同様に考えて

$$\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{n-1}{N}\right) \approx 1 - \frac{n(n-1)}{2N} \quad (N \gg 1)$$

となるので、 n 個の遺伝子のうち初めてどれか二つの遺伝子が共通祖先を持つ時間 T_n は近似的に次の分布を持つ。

$$\text{Prob}[T_n > t] \approx \exp\left[-\frac{n(n-1)t}{2N}\right].$$

以上のことからこの過程は、世代 0 においてサンプルされた n 個の遺伝子の各世代における祖先の数 i を状態変数としてみると、近似的に $i \rightarrow i-1$ への状態遷移率が $i(i-1)/(2N)$ の純粋死滅過程となることがわかる。また中立・任意交配・一定サイズモデルでは、各遺伝子の祖先を過去にさかのぼっていくと、過去にランダムに選ばれた二つの遺伝子が結合されていく確率的に分布する系図ができ上がる。

さて実際の生物では遺伝子が伝わる時に低い確率ではあるが突然変異が起こり、これによってサンプルされた n 個の遺伝子間に遺伝的変異が生じる。上述の確率的な系図の各枝に、枝の長さ t に比例して ut の平均を持つポアソン乱数を与えると、近似的にどのように系図中に突然変異が起こるかを實現することが出来るので、系図の分布とポアソン分布から遺伝的変異の分布を予測することが出来る。例えば集団から 2 個の遺伝子を

取った時に二つの遺伝子間で起こっている突然変異を k 、 n 個の遺伝子を取った時に全系図の中で起こる突然変異の数を S_n で表すとその平均値は

$$E[k] = 2Nu, \quad E[S_n] = a_n(2Nu) \quad (a_n = \sum_{i=1}^{n-1} \frac{1}{i}), \quad (1)$$

となることが知られている。遺伝子が無限個のサイトを持ち、新しい突然変異はこれまでに突然変異が起こったことが無いサイトで起こるとすると（無限サイトモデル）、 k は 2 遺伝子間での異なるサイトの数、 S_n は n 遺伝子の中で多型になっているサイトの数と、実際に観測出来る量に対応する。

Tajima (1989) は (1) の関係を利用して、 n 個の遺伝子配列を得た時に中立・任意交配・一定サイズモデルを検定する次の統計量、Tajima's D 、を提案した。

$$D = \frac{\bar{k} - S_n/a_n}{\sqrt{e_1 S_n + e_2 S_n^2}}. \quad (2)$$

ここで \bar{k} は全ペアの遺伝子についての k の標本平均を表し、分母は分子の標準偏差についての推定値を表す。(1) より $E[\bar{k}] = E[k] = E[S_n/a_n] = 2Nu$ なので分子の期待値はゼロとなり、近似的に D は平均が 0、分散が 1 の分布を持つと考えられる。そこで塩基配列データから得た k, S_n を代入して、 D の値が 0 と有意に異なるかどうかを調べることによって、このモデルを帰無仮説として検定することが出来る。遺伝子系図理論を使ったこのモデルの検定法については他にも多く提案されており、よく使われるものには Fu and Li (1993) のテストや HKA (Hudson et al., 1987) テストなどが有る。

3 サイズが増加する場合

実際の生物集団では、中立・任意交配・一定サイズの仮定のどれかが満たされていない場合も有ると考えられる。例えば Stephens et al (2001) はヒト集団の 82 人のサンプルで 313 遺伝子座を調べると多くの遺伝子座で Tajima's D が負の値を取ったことから、ヒト集団が最近集団サイズの増加を経験したと推測している。そこで集団サイズが指数関数的に増加する時に Tajima's D 等の統計量がどのように分布するかを調べてみた (Sano and Tachida 2005)。現在の時刻を 0 とし、集団サイズは t_e 世代前まで N_0 でありそれ以降次の式に従って増加したとする。

$$N(t) = N_0 \exp[\lambda(t_e - t)/N_0]. \quad (0 \leq t \leq t_e)$$

まず増加率 λ が無限大に近づいたときの極限の系図について調べると、一定サイズ N_0 の集団で得られる系図の終端点にそれぞれ長さ t_e の枝を接合した系図 (rapid-growth-limit) に収束する事がわかった。これは集団サイズが急速に増加してい

る時遺伝子の coalescence が殆ど起こらないことによる。この極限系図で Tajima's D がどのような分布を取るかを調べたところ、平均が負で分散が小さくなることがわかった (Table 1)。また D の平均・分散が t_e の単調減少関数であることも示すことが出来た。次に λ が有限値を取りながら増加した時、この極限にどのように近づくかについてシミュレーションにより調べたところ、 λ の値が 3 を越えると分布の形はかなり極限分布に近づくこともわかった (Table 1)。

Table 1. 集団が増加する時の Tajima's D ($t_e=3$)

| λ | 0 | 0.5 | 1 | 3 | 7 | ∞ |
|-----------------|--------|--------|--------|--------|--------|----------|
| $E[D]$ | -0.041 | -0.520 | -1.170 | -2.065 | -2.222 | -2.278 |
| $\text{Var}[D]$ | 0.901 | 0.574 | 0.300 | 0.085 | 0.062 | 0.054 |

4 集団が分断・融合を繰り返す場合

生物の集団はサイズの変化だけでなく更に複雑な構造を取る事が有る。例えば過去に約 10 万年周期で氷河期と間氷期が繰り返されており、これに伴ってスギ等の樹木やその他の植物集団で分断化・融合が繰り返されたことが花粉化石の分析から推測されている (安田・三好、1998)。このような状況で遺伝的変異量やパターンがどのようなようになるかを調べるために、簡単なモデルを考察した。 t_1 世代続く融合期では集団は単一でサイズは N_1 、 t_2 世代続く分断期ではそれぞれがサイズ N_2 の d 個の分集団に分断され、融合期・分断期が周期的に繰り返されると仮定する。まずこのサイクルが無制限繰り返された後の周期の初め、つまり融合直後に二遺伝子をサンプルした時に、その二つの遺伝子が異なっている確率 θ_∞ を計算した。

$$\theta_\infty = \frac{(\omega_2^{t_2} + (d-1)\nu^{t_2})(1-\omega_1^{t_1})\theta_{1\infty} + (1-\omega_2^{t_2})\theta_{2\infty} + (d-1)(1-\nu^{t_2})}{n - (\omega_2^{t_2} + (d-1)\nu^{t_2})\omega_1^{t_1}}, \quad (3)$$

$$\nu = (1-u)^2, \quad \omega_i = \left(1 - \frac{1}{2N_i}\right)\nu, \quad \theta_{i\infty} = \frac{1-\nu}{1-\omega_i}. \quad (i=1,2)$$

θ_∞ の値を分断期間 t_2 の関数としてみると、 $d=1$ の場合は単調減少関数であるが、 $d \geq 2$ の場合は一旦減少した後増加する関数となる。次に Tajima's D がどのような値を取るかについて調べるために、 $E[S_n]$ を分断化期の集団数 d が 2 の時に数値的に計算したところ $E[k]$ より大きな値となり、 $E[D]$ は集団サイズ増加の場合とは反対に正の値を取る事がわかった。

一般に分化した小集団が融合すると、遺伝子座間の変異の相関を表す連鎖不平衡量が增大することが知られている。そこでこのモデルについても遺伝子座間の相関係数の二乗 ρ^2 を指標にして連鎖不平衡がどのようなになるか調べてみた。その結果、特に d が 2 以

上の小さな値を取る時、集団融合直後の $E[\rho^2]$ が遺伝子座間の組み換え率 r が大きくなってもなかなか減少しないことがわかった。任意交配・一定サイズモデルでの $E[\rho^2]$ の近似値はすでに Hill (1975) によって得られているが、分断・融合モデルでは r の関数として非常に異なった振る舞いを示すので、上記の Tajima's D の振る舞いと合わせて、これらの統計量がこのような集団構造の推定に利用出来ると考えられる。

5 結論

中立・任意交配・一定サイズモデルについては既にかなりその性質が知られているが、これらの仮定のどれかが成り立たない場合についてはまだ研究すべきことが多く残されている。特に集団構造については過去も含めた生物集団についての生態学的知識が蓄積してきており、これらを取り入れたモデリングが可能になっている。初めに述べたように遺伝的変異は様々な進化的要因によって支配されており、それぞれの貢献の程度を明らかにすることが進化機構の理解につながる。ゲノムの時代ではヒトやショウジョウバエなどを含めて大量のデータが蓄積されているので、それを見据えた理論の発展が必要である。

References

- Fu, Y.-X., Li, W.-H., 1993. Statistical tests of neutrality of mutations. *Genetics* 133, 693-709.
- Hill, W., 1975. Linkage disequilibrium among multiple neutral alleles produced by mutation in a finite population. *Theor. Pop. Biol.* 8, 117-126.
- Hudson, R. R., Kreitman, M., Aguade, M., 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116, 153-159.
- Kingman, J. F. C., 1982. The coalescent. *Stoch. Proc. Appl.* 13, 235-248.
- Sano, A., Tachida H., 2005. Gene genealogy and properties of test statistics of neutrality under population growth. *Genetics* (in press).
- Stephens, J. C., Schneider, J. A. *et al.*, 2001. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293, 489-493.
- Tachida H., 2005. Evolution in periodically fragmented populations. (in preparation)
- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis. *Genetics* 123, 585-595.
- Tavaré, S. 1984. Line-of-descent and genealogical processes.. and their applications in population genetic models. *Theor. Pop. Biol.* 26, 119-164.
- 安田喜憲・三好教夫編著, 1998 図説日本植生史. 朝倉書店