

유사종사이의 전이확률에 대한 컴퓨터 계산

길원평(부산대학교 물리학과)

[1] 서론

· 유전자 모형(Genetic Model)

- 진화론자들이 진화를 설명하고자 만든 모형이다.
- 진화론의 개념인, 돌연변이와 자연선택을 고려한다.

· 적합도(fitness)

- 자연 선택을 고려하고자 도입한 양이다. 자연환경에 적응하는 정도를 나타낸다.
- 유전자모형에서는 적합도에 비례하여 새끼를 낳는다고 가정한다.
즉, 적합도가 큰 개체가 많은 새끼를 낳는다고 가정한다.
- 적합도가 유전자서열에 의해만 정해진다고 가정한다.
즉, 유전자서열에 있는 돌연변이의 개수에 따라 적합도가 변한다고 가정한다.
(일반적으로 적합도는 다른 환경요인, 즉 기후, 개체군 내에서의 먹이, 약탈하는 또는 경쟁하는 생물체의 분포 등에 의해서 영향을 받는다.)

· 적합도 경관(fitness landscape)

- 개체군에는 다양한 유전자서열을 가진 개체들이 존재하고, 각 유전자서열에 대응하는 적합도의 값이 주어진다.
- 2차원 평면에 다양한 유전자서열의 적합도를 그림으로 그려서 표현한다면, 산과 골짜기가 있는 경관(landscape)처럼 보인다.

· 유전자서열

- 유전정보를 담은 DNA는 네 개의 염기(A, G, T, C)로 구성되어 있다.
유전자에 있는 염기의 배열을 유전자서열이라 한다.
- 우리가 고려하는 유전자모형에서는 간단히 두 개의 서열원소(+1, -1)로 구성되어 있다고 가정한다.

(예) 유전자서열의 길이 $N=10$ 일 때에, 특정한 유전자서열은 다음과 같다.

(+1,+1,+1,-1,-1,+1,+1,+1,+1,+1)

- 이때에 유전자서열의 종류 수는 2^N 이다.

· 돌연변이의 종류

- ① 치환(substitution) → DNA의 어떤 염기쌍이 다른 염기쌍으로 바뀌는 것
 - ② 삽입(insertion) → DNA에 다른 염기쌍이 첨가되는 것
 - ③ 결실(deletion) → DNA 내의 염기쌍이 제거되는 것
 - ④ 중복(duplication) → DNA의 어떤 부분이 두 배로 되는 것
 - ⑤ 전좌(translocation) → DNA의 어떤 부분이 다른 곳으로 이동하는 것
- < 우리가 사용하는 유전자모형은 '단일염기치환' 만을 고려한다. >

※ 진화론자들이 연구하는 유전자모형을 연구하는 이유

- 진화론자들이 고안한 모형을 사용하여도 진화가 일어날 수 없는 것을 보이려고 함
- 적을 이기려면 적을 알아야 한다.(그 분야의 전문가가 되어야 발언권을 가질 수 있다.)
- 전문 학술지에 투고할 수 있다.(학문 안에서 진화론을 이겨야 한다.)

[2] 유전자모형의 종류

(1) 적응보행모형(Adaptive Walk Model)

- 개체군을 한 점으로 표현한다.
→ 즉, 개체군 내의 유전자서열은 모두 같다고 가정한다.
- 피셔의 정리(Fisher's theorem) → 개체군의 평균 적합도는 항상 증가한다.
- 컴퓨터 계산으로서 확률적으로 약간 다른 유전자서열을 만든다.
→ 새로운 유전자서열의 적합도가 현재보다 커질 때에만 그 유전자변화를 선택한다.
→ 이렇게 하면 적합도가 점점 커지는 방향으로 개체군이 진화한다.
(비유 : 개미가 조금씩 산으로 올라가는 것 같음)
- 적합도가 점점 커져서, 새로운 종으로 진화한다는 것을 보이려는 것이 목적이다.

※ 적응보행모형의 문제점

→ 적합도가 증가하는 방향으로만 진행하기에, 주위보다 적합도가 큰 적합도정상(fitness maximum)에

올라가면 내려올 수가 없다.

(비유 : 개미가 산꼭대기에 올라가면 내려올 수 없다)

→ 조그만 동산(local maximum)에 일단 올라서면, 진짜 높은 산(real maximum)에 갈 수 없다.

종이 바뀌려면, 유전자서열이 굉장히 많이 바뀌어야 하는데, 즉 아주 멀리 떨어진 real maximum까지 가야 하는데 중간의 local maximum에 붙잡혀서 더 이상 변하지 못한다.

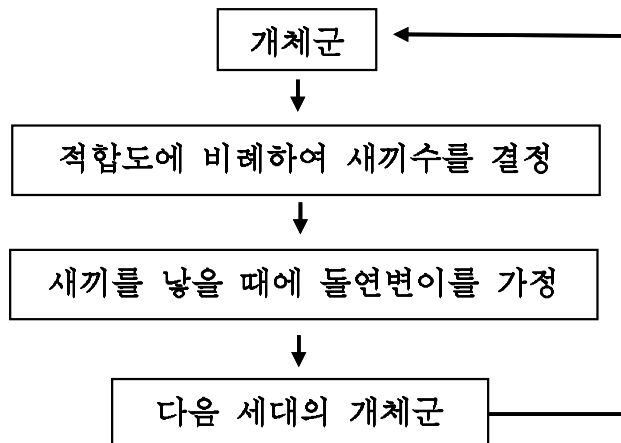
- 따라서 적응보행모형으로는 종이 바뀔 수 없음

(2) 유사종모형 (Quasispecies Model)

· 적응보행모형은 자연선택만 고려했는데, 유사종모형은 자연선택(적합도)과 돌연변이를 모두 고려한다.

- 개체군 내에 다양한 유전자서열이 존재한다고 가정한다.
- 단성생식을 고려 → 성(sex)을 고려하지 않음

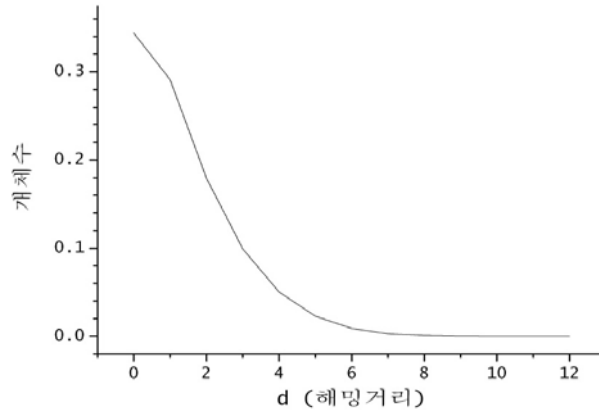
※ 유사종모형의 대략적 과정



※ 유사종(Quasispecies)

→ 오랜 시간동안 새끼 낳는 과정을 반복하면, 결국 정상상태(stationary state)의 유전자 분포를 갖는다.

→ 정상상태의 유전자분포는 아래의 그림과 같이 주서열(master sequence)와 몇 개의 돌연변이를 갖는 유전자서열로 구성된다. 이것을 유사종이라 부른다.



※ 해밍거리(Hamming distance)

→ 주서열로부터 돌연변이를 일으켜 달라진 갯수를 해밍거리라고 한다.

(예) 어떤 유전자서열이 주서열과 2개가 다르면, 주서열과 유전자서열의 해밍거리는 2이다.

※ 적합도 경관의 종류

① 단순경관 (simple landscape)

주서열만 특별히 큰 적합도($w = e^{kN}$)을 갖고, 나머지 서열은 모두 낮은 적합도($w = 1$)를 갖는다.

② 배수경관 (multiplicative landscape)

유전자서열의 적합도는 주서열로부터 해밍거리, 즉 돌연변이를 일으킨 서열원소의 갯수 d 에 의해서

다음 식으로 주어진다.

$$w = \exp[k(N - d)]$$

③ 계단경관 (truncated landscape)

단순경관과 배수경관의 중간형태로서, 주서열과 주서열에 가까운 서열, 즉 해밍거리가 작은 서열은

큰 적합도를 갖고, 나머지 서열은 작은 적합도($w = 1$)를 갖는다.

$$w = \exp[k(N - d)] \quad \text{if } d < L - 1$$

· 앞으로 위 식에서의 k 를 적합도멧음변수라고 부르자.

[3] 유사종모형의 결과

(1) 오차문턱(error threshold)

- 돌연변이율(μ)이 어떤 특정한 값보다 커지면, 유전자분포가 유사종의 형태를 유지하지 못하고, 무질서한 상태가 된다. → 즉 유전정보를 모두 잃어버린다.
- 유전정보를 모두 잃어버리게 되는 돌연변이율을 오차문턱이라고 부른다.
- 단순경관에서 적합도를 고정시키고 오차문턱과 유전자서열길이(N)와의 관계를 보면 대략 다음과 같다.

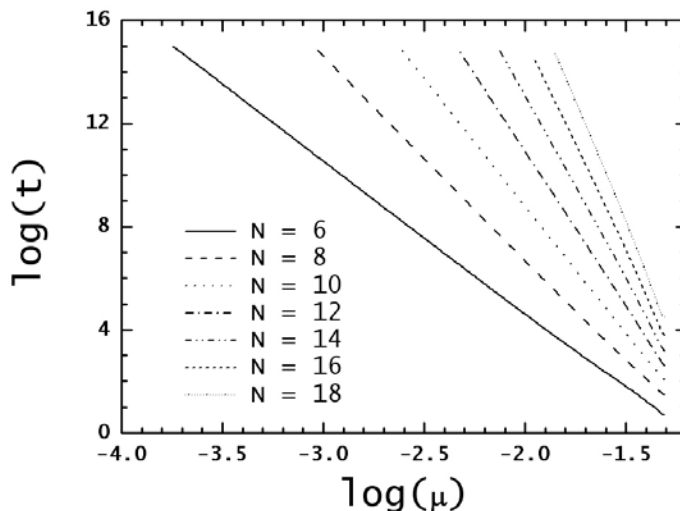
$$\mu_c \approx \frac{1}{N}$$

- ※ 오차문턱의 존재가 진화론자를 당황하게 한다.
 - 유전자서열의 길이가 증가할수록, 더 작은 돌연변이율에서 오차문턱이 생기고 유전정보를 잃게 된다.
 - 따라서 오차문턱은 긴 유전자서열길이를 가진 생물체로 진화할 수가 없게 만든다. 그리고 실제 생물체는 매우 긴 유전자서열을 갖고 있다.

(2) 유사종 사이의 전이확률

1) 단순경관

- 주서열 근처에서만 높은 적합도를 갖게 하고 정상상태의 유전자분포에 도달하게 한 후에, 갑자기 주서열과 정반대되는 서열에도, 즉 모두 돌연변이를 일으킨 서열에도 높은 적합도를 주고, 반대편 서열로 전이할 확률을 계산하였다.
- 컴퓨터로 원래의 유전자분포에서 반대편 유전자서열을 가진 분포로 바뀔 시간, 즉 통과시간을 구하였다. 통과시간은 전이확률과 반비례의 관계를 가진다.
- 아래 그림은 여러 유전자서열길이(N)에서 적합도멱음변수(k)를 0.05로 고정시키고 돌연변이율(μ)을 변화시켜 가면서 통과시간(t)을 계산하였다.



위의 결과로부터 돌연변이율이 감소함에 따라 통과시간이 증가하는 것을 볼 수 있다. 그리고 유전자서열길이(N)이 증가함에 따라 증가속도가 더 크다는 것을 볼 수 있다. 위의 그림에서 통과시간과 돌연변이율이 모두 로그눈금으로 그려져 있고, 그래프가 거의 직선이기 에, $\log(t) \propto s \log(\mu)$ 의 관계이다. 그리고 직선의 기울기 s 는 대략 $-N$ 으로 주어진다.

다. 따라서 $t \propto (\frac{1}{\mu})^N$ 이다. 이 결과로부터 유전자서열길이 N 이 증가함에 따라, 돌연변이율 μ 이 감소함에 따라 통과시간 t 가 급격히 증가한다는 것을 알 수 있다.

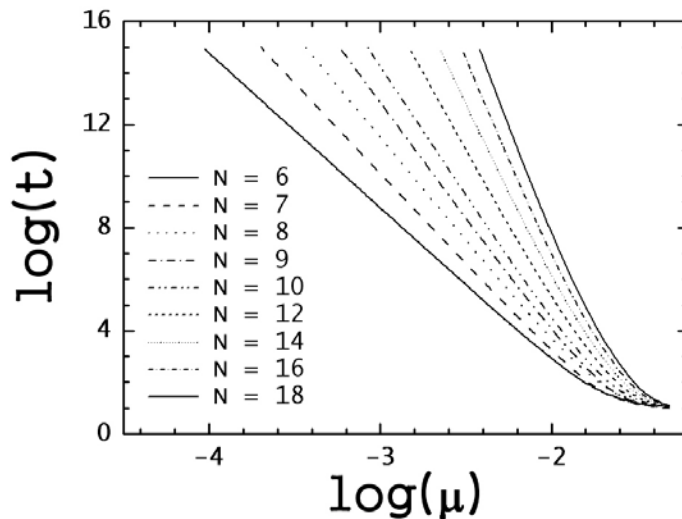
2) 배수경관

· 배수경관에서 정상상태의 유전자분포에 도달하게 한 후에, 갑자기 해밍거리 d 의 적합도를 해밍거리

$N-d$ 의 적합도와 같도록 하였다. 이와 같은 대칭배수경관에서 돌연변이가 전체 유전자서열원소의

절반 이상 일으킨 개체수가 총 개체수의 1% 가 되는데 걸리는 시간을 컴퓨터로 계산하였다.

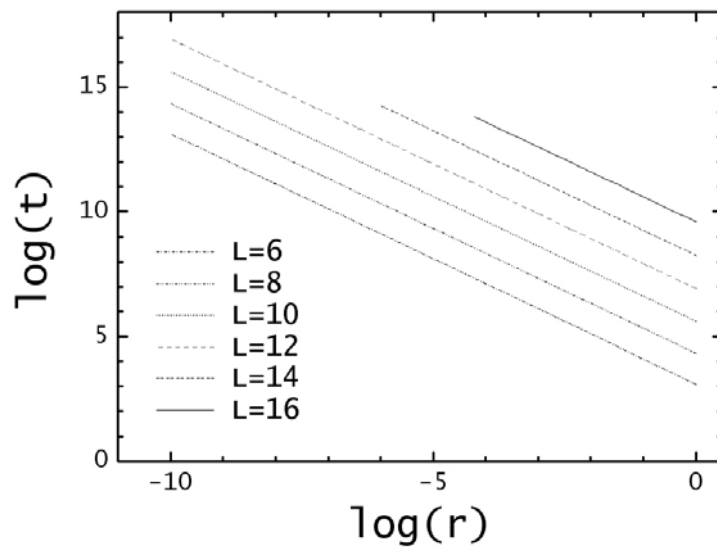
- 아래 그림은 여러 유전자서열길이(N)에서 적합도멧음변수(k)를 0.05로 고정시키고 돌연변이율(μ)을 변화시켜 가면서 통과시간(t)을 계산하였다.



위의 결과로부터 배수경관에서의 통과시간은 단순경관의 통과시간과 비슷한 경향을 가진다는 것을 알 수 있다. 즉 돌연변이율이 감소함에 따라 통과시간이 증가하고, 유전자서열길이 증가함에 따라 통과시간의 증가속도가 커진다는 것을 볼 수 있다. 단순경관에서도 같이 통과시간과 돌연변이율이 모두 로그눈금으로 그려져 있고, 그래프가 거의 직선이고, 직선의 기울기가 대략 $-N$ 이기에, $t \propto (\frac{1}{\mu})^N$ 이다. 이 결과로부터 단순경관과 배수경관에서 유전자서열길이 N 이 증가함에 따라, 돌연변이율 μ 이 감소함에 따라 통과시간 t 가 급격히 증가한다는 것을 볼 수 있었다.

3) 세대교체를

위의 두 결과는 세대교체율 r 을 1(=100%) 로 놓고 계산하였다. 즉 모든 부모들은 새끼를 낳고 죽는 것으로 가정하였다. 이제는 세대교체율을 변화시켜 가면서 통과시간이 어떻게 변화하는지 살펴보겠다. 아래 그림은 단순경관에서 여러 유전자서열길이(N)에서 적합도멧음변수(k)를 0.05로, 돌연변이율(μ)은 0.01로 고정시키고 세대교체율을 변화시켜 가면서 통과시간(t)을 컴퓨터로 계산하였다.



그 결과 세대교체율이 감소함에 따라 통과시간이 증가함을 볼 수 있었다. 즉 통과시간과 세대교체율이 모두 로그눈금으로 그려져 있고, 그래프가 거의 직선이고, 직선의 기울기가 -1 이기에, 세대교체율과 통과시간은 반비례의 관계를 갖고 있음을 알 수 있다. 또한 컴퓨터 계산으로 배수경관도 같은 경향을 가지고 있음을 알 수 있었고, 세대교체율은 상한선이 존재함을 보았다.

· 계산한 결과, 통과시간은 돌연변이율이 감소할수록, 유전자서열길이가 증가할수록 급격히 증가한다는

것을 보았다. 예를 들어서 돌연변이율 $\mu = 10^{-6}$ 이고, 유전자서열길이 $N = 10^9$ 이라면, 통과시간은

$t \propto 10^{6 \times 10^9}$ 이 된다. 이 값은 천문학적인 숫자로서 사실상 통과가 불가능하다는 것을 알 수 있다.

[4] 요약

진화를 설명하려고 고안한 유사종모형을 사용하여 유사종 사이의 전이확률을 컴퓨터로 계산하였다. 적합도의 분포로서 단순경관과 배수경관을 고려하고, 유사종모형의 변수를 변화시키며 전이확률을 계산하였다. 그 결과, 통과시간이 유전자서열길이가 증가함에 따라 지수적으로 증가하는 것을 볼 수 있었다. 즉 전이확률이 유전자서열길이가 증가함에 따라 지수적으로 감소한다는 것을 볼 수 있었다. 또한 통과시간이 세대교체율과는 반비례의 관계를 갖는다는 것을 보았고, 세대교체율에는 상한선이 존재한다는 것을 보았다.

따라서 유사종모형에서 실제적인 돌연변이율을 고려하여서 많은 돌연변이를 요구하는 전이확률, 즉 통과시간을 계산한다면, 엄청난 통과시간이 요구되기에 많은 유전자차이를 갖는 유사종 사이의 전이는 사실상 불가능하다는 것을 볼 수 있었다.