

진화시간에 대한 컴퓨터 시뮬레이션 결과

이규상

부산대학교 자연과학대학원 물리학과

요약

세상은 진화론을 진리인양 가르치고 있다. 그로 인해 많은 그리스도인들이 하나님의 창조섭리에 대한 혼란을 느끼고 있다. 그러므로 진화론이 과연 신뢰할만한 과학이론인지를 그리스도인으로써 연구하여 하나님의 창조섭리를 나타낼 필요가 있다. 본 연구는 진화 과정을 수식화하고, 컴퓨터 시뮬레이션을 통해 진화시간, 즉 진화가가능성에 대해 연구하였다.

이전의 연구들은 유전자서열요소의 허용 상태 수가 두 개인, 두 상태 돌연변이-자연선택 모델에서 광범위하게 연구하였다. 본 연구는 무한한 개체수에 대해 n 상태 반수체 결합 불연속시간 돌연변이-자연선택(HCDMS) 모델에서의 오차문턱을 연구하였다. n 상태 HCDMS 모델에서 단순지형일 때의 오차문턱은 $k \ll 1$ 일 때 $k/(n-1)$ 로 표현됨을 발견하였다. 여기서 k 는 적응도 맷음변수이다. 세, 네 상태 모델 비대칭단순경관일 때의 대립유전자들은 이차원 그룹으로 분류하여 다양한 동적 성질에 대하여 계산하였다. 모든 유전자서열요소 사이의 돌연변이율은 같다고 가정하면, $s > 0$ 일 때의 진화시간은 적응도맷음변수를 증가시키면 큰 적응도맷음변수에서 유한한 값에 수렴하였다. 여기서 s 는 선택이익이다. $s = 0$ 일 때의 적응도맷음변수가 증가하면, 화시간은 지속적으로 증가하였다. $s < 0$ 일 때의 진화시간은 적응도맷음변수가 증가하여서 임계 적응도맷음변수에 접근하면 발산하였다. 컴퓨터 시뮬레이션의 결과, 확장맷음변수 E 를 고정시킨 상태에서 $s > 0$ 이면 진화시간은 긴 유전자서열을 가질 때 포화되었고, $s = 0$ 이면 유전자서열길이가 증가할 때 진화시간이 지속적으로 증가하였다. 네 상태 모델에서 두 유전자서열요소 사이의 돌연변이율이 각각 다르다고 가정하고, 돌연변이율의 비율 $r_{ij} = \mu_{ij}/\mu_{14}$ 이라고 놓았다. 네 상태 모델에서 $s = 0$, $r_{23} = r_{42} = 1$ 일 때, 진화시간은 $r_{12} > 1$ 이면 적응도맷음변수가 클 때에 유한한 값에 수렴하였고, $r_{12} = 1$ 일 때는 진화시간이 지속적으로 증가하였고, $r_{12} < 1$ 일 때는 적응도맷음변수가 임계적응도맷음변수에 접근함에 따라 발산하였다.

즉, 본 연구의 일반적인 조건에서는 진화시간이 유전자서열길이 증가에 따라 급격하게 증가함으로써 진화론에서 주장하는 충분히 긴 시간이 있다하더라도 진화가 이루어질 수 없다는 것을 보였다. 진화론이라는 것이 그럴듯해 보여도 결국에는 진화론에서 말하는 충분히 긴 시간이라는 것이 실제로는 진화를 일으킬 만큼의 충분히 긴 시간이 아니다. 결국에는 이 세상을 창조한 하나님의 창조섭리 없이는 지금의 인간은 있을 수 없다.

I. 서론

완전하고 무오한 성경에 적혀있는 것처럼 이 세상은 하나님께서 하나님의 질서와 창조섭리대로 창조하셨다. 그리고 하나님은 인간을 하나님의 형상대로 존귀하게 지으시고 이 세상을 다스리게 하셨다. 그리고 우리 그리스도인들은 이것을 믿고 이 세상 가운데 하나님의 형상대로 살려고 노력하는 사람들이다. 그러나 이 세상이 가르치고 말하는 자연의 창조섭리는 실히 증명되지도 될 수도 없는 진화론이며 이것을 어린 학생들에게 진리인양 믿게 한다. 이로 인해 어릴 때부터 진화론을 비판적 수용 능력 없이 받아들인 학생들은 교회 안에서도 세상이 말하는 진리와 성경의 진리 가운데서 흔들린다. 이는 단순히 볼 문제가 아니라 매우 심각한 문제다. 세상에서는 점점 더 경험하지 못하고, 눈에 보이지 않고, 증거가 되지 않는 것은 없는 것과 마찬가지로 가르치는 과학신봉이 심해진다. 이러한 분위기 속에서 자라난 많은 학생들과 어설프게 진화론을 접한 많은 이들은, 진화론의 허구에 대해서는 생각해볼 겨를도 없이 단순히 진화론이 생명의 기원에 대해 설명하는 유일한 과학적 가설이라는 이유만으로, 그것이 이 세상의 진리인 줄 알고 살아가며 성경에 있는 하나님의 창조섭리를 배척하고 살아가게 될 것이다.

우리는 이러한 세상 가운데 과학을 탐구하는 하나님의 종으로써, 진화론이 과연 세상이 말하는 대로 진리로 받아들일만한 것인지 그리고 진화론이 틀렸다면 왜 틀렸는지를 제시해주어야 한다고 생각한다. 이러한 노력이 믿음을 지키기 어려운 시대를 살아가고 있는 현대의 그리스도인들에게 우리의 믿음이 결코 틀리지 않았음을 알게 하는데 도움이 될 것이다. 또한 진화론을 잘 모르면서 신봉하고 있는 불신자들에게도 논리적인 근거 위에서 바른 판단을 하는데 큰 도움이 될 것이다. 부족한 인간이 하나님의 놀라운 창조섭리를 모두 알 수는 없겠지만, 이 세상 모든 만물은 하나님의 창조섭리대로 지어졌기에 모든 학문에서 하나님의 진리를 드러내어서 그것을 사람들이 알도록 하는 그리스도인 학자들의 역할이 중요하다.

‘이기적 유전자’의 작가로 유명한 리처드 도킨스의 ‘눈먼 시계공’이라는 책은 신학자들의 진화론을 비판하고 창조론을 설명한 논문들에 대해 비판하면서 진화론을 설명하고 있다. 그러나 리처드 도킨스 또한 진화론이 갖는 오류에 대해 자유롭지 못한 것을 알 수 있다. 리처드 도킨스는 신학자가 진화론은 원숭이가 타자기를 무작위로 쳐서 셰익스피어의 작품을 만들어내는 것과 같다고 비판한 것을 비판하며 진화론을 누적선택을 사용하여 설명하는데 이 설명은 저자의 설명을 따라가다보면 그럴듯해 보이지만 사실 설정 자체에 두 가지 오류가 있다. 첫 번째, 목표로 한 돌연변이가 일어나면 다시 역방향으로 돌연변이는 일어나지 않는다고 설정하였는데 실제 돌연변이는 목표 따위가 없기 때문에 역방향 돌연변이도 일어난다. 도킨스는 이 사실을 무시하고 시뮬레이션하여 원숭이가 타자를 무작위로 쳐서 셰익스피어의 햄릿에 나오는 한 문장을 얻어내는데 40세대 정도면 충분하다고 하였지만 실제 역방향 돌연변이까지 고려하면 엄청난 차이가 생긴다. 그리고 실제와 같이 유전자서열길이 길어지면 이 차이는 더 큰 차이를 불러온다. 두 번째, 각 세대마다 목표로 한 것과 가장 유사한 결과만을 인위적으로 취하여 다음 세대를 번식시켰지만 이는 실제 유사종분포를 무시한 것이다. 실제 돌연변이도 적응도라는 것을 갖지만 그것은 분포도를 가지고 있기 때문에 목표로 한 값만 살아남고 나머지는 없어지는 도킨스의 방식과는 전혀 다른 것이다. 본 논문에서는 이런 잘못 된 설정을 바로잡아서 시뮬레이션하였고, 실제 진화가 이루질 수 있는가를 연구하였다.

본 연구는 진화이론학자들이 진화를 설명하기 위하여 고안된 모델로서 돌연변이와 자연선택만을 고려하는 돌연변이-자연선택모델(mutation-selection model)을 사용하고자 한다. 많은 과학자들은 자연선택(natural selection), 돌연변이(mutation), 이주(migration), 유전적 떠돌음(genetic drift), 무작위적 환경변화(random environment), 재조합(recombination) 등과 같은 이때까지 제안된 여러 진화적 요인 중에서 자연선택과 돌연변이는 개체군의 빈도분포를 변화시키는 두 가지 주요 진화 요소로 간주하고 있다.[1-3] 돌연변이와 자연선택이 중요한 진화론적 요인으로 간주되는 이유는 돌연변이가 유전적 변화를 일으키는 필수 불가결한 요소이며, 자연선택은 살아있는 유기체의 기본 특성 중 하나를 기술하기 때문이다. Eigen은 생체고분자물질(biopolymers)의 복제과정을 설명하기 위해 돌연변이와 자연선택만을 고려한 진화 모델을 제안했다.[4] Eigen의 제안 후에 돌연변이-자연선택 모델은 많은 통계 기법을 적용할 수 있는 단순성 때문에 매우 상세하게 연구되었다.[5-10] Eigen모델에 의해 열게 되는 중요한 특성 중의 하나는 단순지형에서 오차문턱이 존재한다는 것인데, 이 단순지형은 가장 단순하기에 가장 많이 연구된 적응도지형이다.[11] 만약 돌연변이율이 오차문턱이라고 하는 임계값보다 작으면, 개체군이 최적의 대립유전자 주위에 분포한다. 이러한 국소화된 개체분포를 유사종이라고 부른다. 만약 돌연변이율이 오차문턱보다 크면, 개체군이 모든 대립유전자에 골고루 분포하여 비국소화된 상태가 되어서 결국 유전정보를 잃어버리는 결과를 낳는다. 반수체 돌연변이-자연선택 모델은 단순지형의 오차문턱에서 개체군의 분포가 국소화된 상태 모델에서 비국소화된 상태 모델로 급격히 변화하는 것을 보여준다.

돌연변이-자연선택 모델의 동적, 평형적 특성은 지난 40년 동안 광범위하게 연구되어 왔다. 이 모델은 돌연변이 기제(mechanism)에 따라 크게 둘로 나누어진다. 생식과정에서의 복제 오류에 의해서 생기는 돌연변이를 고려하는 모델을 결합 돌연변이-자연선택모델(coupled mutation-selection model)이라고 부르며,[12-22] 자외선, 화학물질 등의 돌연변이원(mutagen)에 의해서 일생 중 어느 때라도 가능한 돌연변이를 고려하는 모델을 분리 돌연변이-자연선택모델(decoupled mutation-selection model)이라고 부른다.[23-25]

실제 DNA 유전자서열에서는 네 가지 상태를 선택할 수 있음에도 불구하고, 위에서 언급한 대부분의 연구에서는 유전자서열요소가 두 가지 상태를 선택할 수 있다고 가정한 두 상태 모델을 고려하였다. 두 상태 모델은 네 종류의 염기를 퓨린과 피리미딘의 두 가지 유형으로 나눔으로써 DNA를 단순화시켜 설명하였다고 볼 수 있다. 그러나, 실제 DNA에 의해 나타나는 일부 현상은 이렇게 단순화된 두 상태 돌연변이-자연선택 모델을 사용해서는 설명되지 않는다. 대립유전자 종류의 총 수가 4^L 에서 2^L 로 감소하기 때문에 DNA를 단순화한 두 상태 모델에서 표현형의 총 수는 기존의 다양한 표현형을 통합함으로써 감소될 것으로 유추된다. 여기서 L은 대립유전자의 서열길이이다.

본 논문은 컴퓨터 시뮬레이션을 사용하여 실제 DNA와 같이 유전자서열요소가 네 가지 상태를 선택할 수 있다고 가정한 네 상태 반수체 결합 불연속시간 돌연변이-자연선택(HCDMS) 모델을 무한한 개체군인 경우에 조사했다. 적응도지형을 단순지형에서 비대칭단순지형으로 갑자기 바꾼 후에 상대밀도의 시간적 변화와 진화시간과 같은 여러 특성을 조사했다. 이 연구에서 적응도지형의 급격한 변화는 환경변화에 따른 새로운 최적의 대립유전자 출현을 설명한다고 볼 수 있다. 본 연구에서 진화시간은 상당수의 개체가 새로운 최적의 대립 유전자에 도달하기 위해 적응

계곡을 통과하는데 걸리는 시간 간격을 의미한다. 그러므로 진화시간은 새 특성을 나타낼 수 있는 새로운 최적의 대립유전자로의 진화시간과 동일하다고 볼 수 있다. 모델 변수들을 변화시켜가면서 진화시간에 대한 컴퓨터 시뮬레이션을 하면, 적응계곡을 통과하는 데 중요한 기제 (mechanism)에 대한 정보를 알 수 있으므로, 진화시간에 대한 시뮬레이션은 과학자들에게 매력적이라고 볼 수 있다. 상대밀도의 시간적 변화와 진화시간을 네 상태 HCDMS 모델에서 계산한 후에 두, 세 상태 HCDMS 모델에서의 상대밀도의 시간적 변화와 진화시간과도 비교해 보았다.

II. 여러 상태 결합 불연속시간 돌연변이-자연선택 모델

본 논문은 유전자서열요소의 허용 상태 개수가 n 개인 반수체 결합 불연속시간 돌연변이-자연선택 (HCDMS; haploid coupled discrete-time mutation-selection) 모델을 사용한다. 이 모델은 무성생식으로 번식하고 시간이 세대별로 불연속적이라고 가정한다. 본 논문에서는 개체 수는 무한하고 점 돌연변이만을 고려하였다. 대립유전자(allele) S_i 를 길이 L 의 유전자서열 $S_i = (s_1^{(i)}, s_2^{(i)}, \dots, s_L^{(i)})$ 로 나타낸다. 유전자서열요소는 $+1, 1 - \frac{2}{n-1}, 1 - \frac{4}{n-1}, \dots, -1$ 과 같이 $+1$ 과 -1 사이의 n 개의 등간격 값으로 나타낼 수 있다. 이때 대립유전자의 종류는 n^L 이다. 두 대립유전자 간의 서로 다른 서열요소 개수를 해밍거리(Hamming distance)로 표현된다. 대립유전자 S_i 의 상대밀도 $x_i = (n_i/N)$ 는 다음 방정식에 의해 시간에 따라 변한다. 여기서 n_i 는 대립유전자 S_i 를 가지는 개체수이며, N 은 총 개체수이다.

$$x_j(t+1) = \frac{\sum_i Q_{ji} w_i x_i(t)}{\bar{w}} \quad (1)$$

$$\bar{w} = \sum_i w_i x_i(t) \quad (2)$$

위 식에서 집단의 평균 적응도는 \bar{w} 로 표시된다. 본 논문에서는 두 서열요소 사이의 돌연변이율이 같은 경우와 다른 경우를 모두 고려하였다. 모든 두 서열요소 사이의 돌연변이율이 같은 경우에는, S_i 에서 S_j 로 변하는 돌연변이 확률은 $Q_{ji} = (1 - \mu_t)^{L - d_{ij}} \mu_t^{d_{ij}}$ 로 표현된다. 여기서 μ 는 세대 당 두 서열요소 사이의 돌연변이율을 의미하고, μ_t 는 어떤 상태에서 $(n-1)$ 개의 다른 상태로의 총 돌연변이율을 의미하고, d_{ij} 는 S_j 와 S_i 사이의 해밍거리를 의미한다. 그러므로 총 돌연변이율, μ_t 는 $(n-1)\mu$ 와 같다.

단순지형이란 최적의 대립 유전자 $S_0 = (1, 1, \dots, 1)$ 가 최대 적응도 1을 가지며, 나머지 모든 대립유전자는 $e^{(-kL)}$ 의 적응도를 갖는 것이다. 여기서 k 는 적응도뺏음변수(fitness parameter)라고 한다. 처음에는 모든 개체가 최적의 대립유전자를 갖는다고 가정한다. 식 (1)에 의해 대립유전자 분포를 변화시키면, 시간이 충분히 흐른 후에는 집단의 분포가 정상상태(stationary state)가 된다.

정상상태가 된 후에 단순지형에서 아래에 정의된 비대칭단순지형으로 갑자기 바꾼다.

$$w = \begin{cases} 1, & \text{if } S = S_0, \\ 1 + s, & \text{if } S = S_0^*, \\ e^{(-kL)}, & \text{otherwise,} \end{cases} \quad (3)$$

최고의 적응도는 최적의 대립유전자인 S_0 와 최적의 대립유전자의 정반대인 대립유전자 $S_0^* = (-1, -1, \dots, -1)$ 에 주어진다. 그러므로 이 두 최고의 적응도는 적응도 계곡에 의해 분리되어 있다. 이 적응도 계곡의 깊이는 적응도뺏음변수 k 에 의해 결정된다. 그리고 변수 s 를 선택이익(selective advantage)이라고 부른다.

네 상태 모델에서는 두 서열요소 사이의 돌연변이율을 각각 다르게 놓는다. 네 상태 모델에서

는 최고 적응도가 최적대립유전자 S_0 에 할당되고, 그 반대인 대립 유전자 $S_0^* = (4, 4, \dots, 4)$ 를 이 연구에서 역전대립유전자라고 부른다. 여기서의 4는 앞에서의 -1이라고 생각해도 좋다. 네 상태 모델에서는 유전자서열요소의 값이 (1, 1/2, -1/2, -1)로 이루어지고, 편의상 이를 (1, 2, 3, 4)로 표기한다.

두 서열요소 i 와 j 사이의 돌연변이율을 μ_{ij} 로 표현하고, $\mu_{ij} = \mu_{ji}$ 라고 가정한다. 또한 본 연구에서는 $\mu_{12} = \mu_{13}$, $\mu_{42} = \mu_{43}$ 으로 가정함으로써, 서열 요소 '2'와 '3'이 동일한 돌연변이율을 가진다.

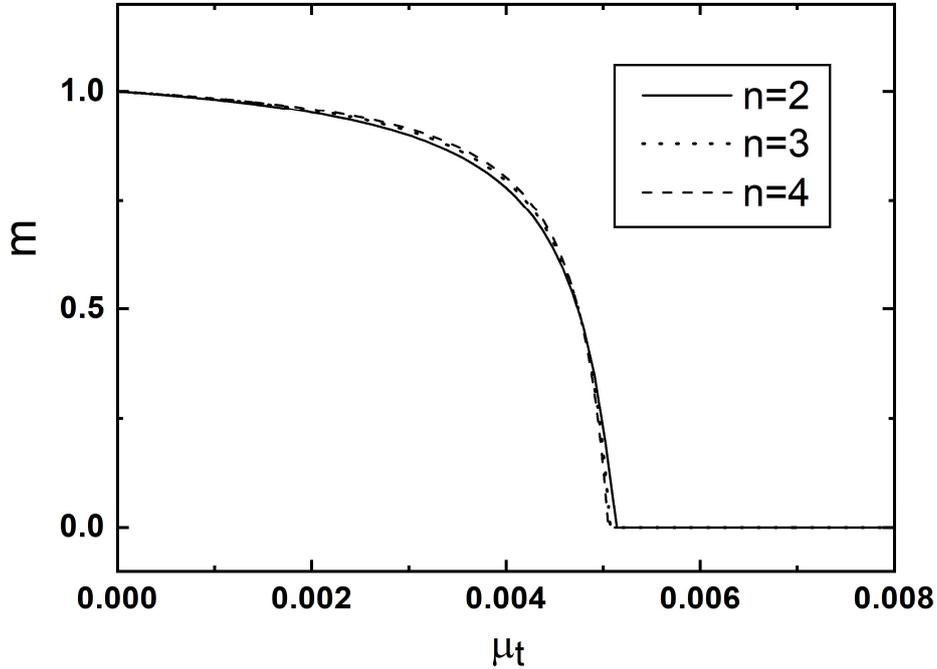
III. 결과

이 장에서는 무한한 개체 수인 유전자서열 요소의 허용 상태 수가 n 개인 경우의 반수체 결합 불연속시간 돌연변이-자연선택(HCDMS) 모델의 결과를 제시한다. 비대칭 단순지형으로 바뀐 시간을 $t = 0$ 으로 설정하였다. 단순지형에서 정상 상태의 질서맺음변수 m 은 아래와 같이 정의된다.

$$m = \frac{1}{L} \sum_{j=1}^L \langle s_j \rangle \quad (4)$$

$$\langle s_j \rangle = \sum_{i=1}^{2^L} x_i s_j^{(i)} \quad (5)$$

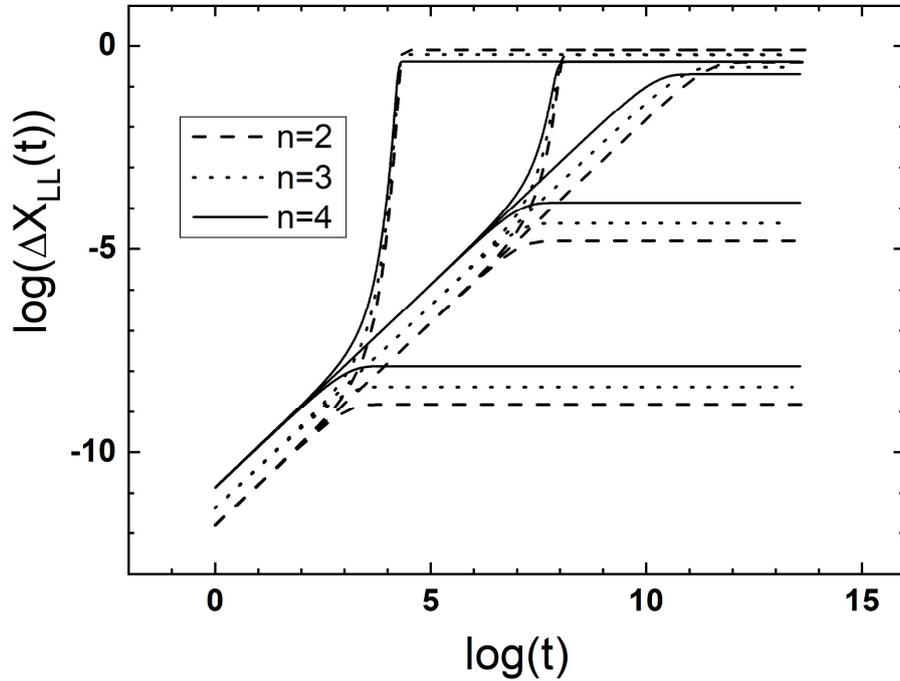
위 식에서 $s_j^{(i)}$ 는 유전자서열 S_i 의 j 번째 서열요소를 뜻하며, x_i 는 개체군 내에서 유전자서열 S_i 의 상대밀도이다. 식 (5)에서 모든 개체가 최적의 대립 유전자를 가지는 질서 정연한 상태에서는 질서맺음변수 $m = 1$ 이 된다. 다른 한편으로, 집단의 모든 개체가 가능한 모든 대립 유전자에 골고루 분포하는 무질서 상태에서는 질서맺음변수 $m = 0$ 이 된다. 모든 서열요소사이의 돌연변이율이 동일하다고 가정하는 것을 Jukes-Cantor 돌연변이체계라고 부른다.



[그림 1] 두, 세, 네 상태 HCDMS 모델에서 단순지형이며 $L = 20, k = 0.005$ 일 때 총 돌연변이율 $\mu_t (= (n-1)\mu)$ 에 따른 정상상태의 질서맺음변수(m) 변화

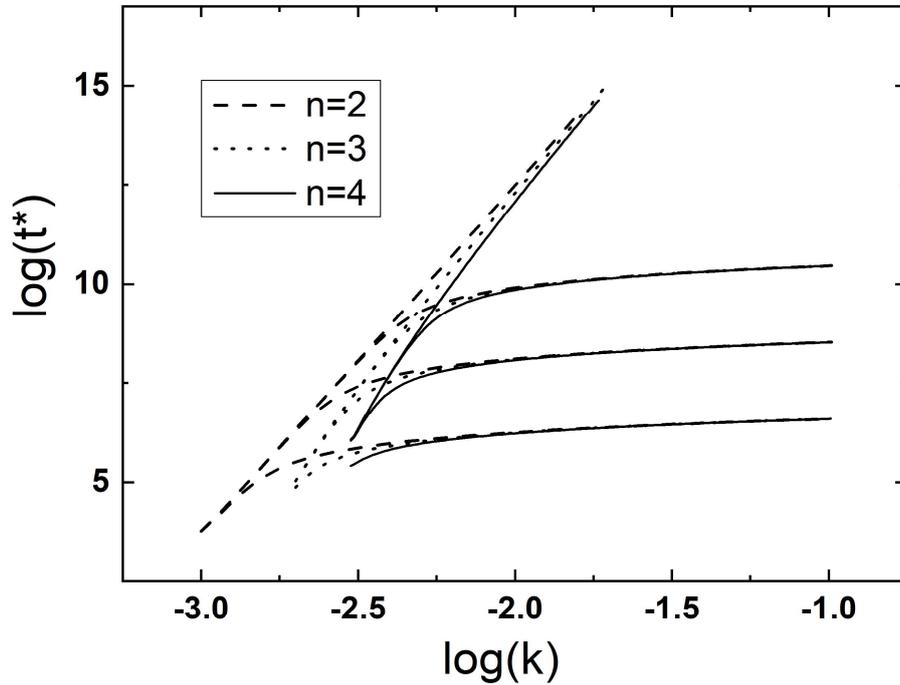
[그림 1]은 Jukes-Cantor 돌연변이체계를 가진 두, 세, 네 상태 HCDMS 모델에서 유전자서열길이 $L = 20$, 적응도땀음변수 $k = 0.005$ 일 때, 총 돌연변이율 $\mu_t (= (n-1)\mu)$ 의 함수로서 단순지형의 정상상태에서의 질서맺음변수 m 을 보여준다. [그림 1]에서 모든 질서맺음변수는 총 돌연변이율 μ_t 가 적응도땀음변수 $k = 0.005$ 와 비슷해질 때에 0으로 급격히 감소했다. 추가 시뮬레이션은 서열 길이가 증가할 때 질서맺음변수의 값과 유사한 총 돌연변이율 근처에서 질서맺음변수가 더 빠르게 0으로 감소함을 보여준다. 두 상태 모델에서 단순지형일 때 임계돌연변이율 μ_c 또는 오차문턱의 이론적인 값은 $1 - e^{-k} (\approx k, \text{ if } k \ll 1)$ 으로 주어진다. 따라서 [그림 1]의 결과로부터 단순지형에서 n -상태 모델의 오차문턱은 $k \ll 1$ 일 때 $k/(n-1)$ 로 주어짐을 알 수 있다.

무한한 개체 수일 때 유전자서열요소의 허용 상태 수가 네 개인 네 상태 HCDMS 모델에서는 $\mu_{ij} = \mu_{ji}, \mu_{12} = \mu_{13}, \mu_{42} = \mu_{43}$ 으로 가정했기 때문에 $\mu_{12}, \mu_{14}, \mu_{23}, \mu_{42}$ 와 같은 네 개의 돌연변이율만 독립적이다. 컴퓨터 시뮬레이션을 한 결과, 오차문턱으로 인하여 네 상태 모델에서는 $k \geq 2\mu_{12} + \mu_{14}$ 이란 제한이 생김을 볼 수 있었다. 돌연변이율 μ_{14} 는 두, 세 상태 모델에서 돌연변이율 μ 와 동일한 값으로 설정되었고, 돌연변이율의 상대 비율 r_{ij} 는 μ_{ij}/μ_{14} 로 정의되었다.



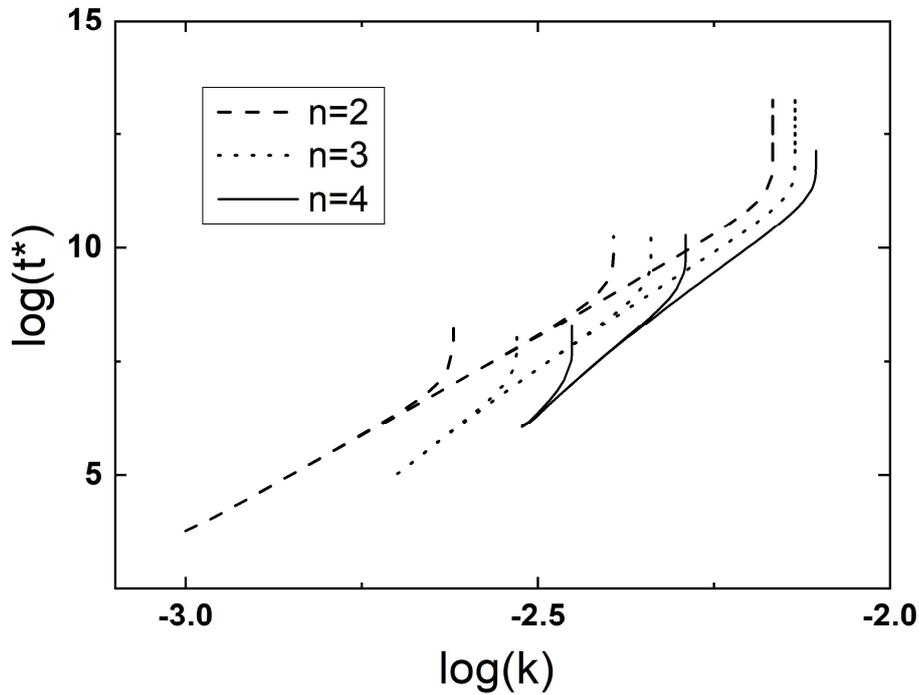
[그림 2] 두, 세, 네 상태 HCDMS 모델에서 $L = 10, k = 0.005, \mu = 0.001$ 일 때 시간의 로그값 $\log(t)$ 에 따른 역전대립유전자의 상대밀도 증가의 로그값 $\log(\Delta X_{LL}(t))$ 변화. Jukes-Cantor 돌연변이 체계에서 $\log(\Delta X_{LL}(t))$ 의 컴퓨터 시뮬레이션 결과는 왼쪽에서 오른쪽으로 선택이익, $s = 10^{-3}, 10^{-7}, 0, -10^{-7}, -10^{-3}$ 에 대한 것을 나타내며, 유전자서열요소의 허용 상태 수가 두 개인 경우($n=2$)는 대시로, 세 개인 경우($n=3$)은 점선으로, 네 개인 경우($n=4$)는 실선으로 그려져 있다.

[그림 2]는 Jukes-Cantor 돌연변이 체계이고 두, 세, 네 상태 HCDMS 모델에서 $L = 10, k = 0.005, \mu = 0.001$ 일 때 시간의 로그값 $\log(t)$ 에 따른 역전대립유전자의 상대밀도 증가의 로그값 $\log(\Delta X_{LL}(t))$ 의 변화를 보여준다. [그림 2]에서 $\Delta X_{LL}(t)$ 는 특성 시간 τ 까지 선형적인 시간 의존성을 가지며, 특성시간의 크기는 대략 선택이익 절댓값의 역수 $1/|s|$ 이다. 또한 $s > 0$ 일 때 $\Delta X_{LL}(t)$ 은 시간에 따라 지수적으로 증가하고, 증가비율은 대략 s 의 크기와 같다. $s < 0$ 일 때 $\Delta X_{LL}(t)$ 가 특성 시간 후에 포화된다. 흥미롭게도 두, 세, 네 상태 모델에서 $\Delta X_{LL}(t)$ 은 특성시간 $\tau (\approx 1/|s|)$ 까지의 선형적 시간 의존성, 특성시간 후에는 $s > 0$ 에서의 지수적 시간 의존성, $s < 0$ 에서의 포화 등과 같은 공통 성질을 갖는다. [그림 2]에서 $s < 0$ 이고 시간이 무한대일 때 포화된 역전대립유전자의 상대밀도 증가 값 $\Delta X_{LL}(t = \infty)$ 이 네 상태 모델에서 가장 크며, 두 상태 모델에서 가장 작고 세 상태 모델에서 중간임을 볼 수 있다. 이는 시간 $t = 1$ 일 때 역전대립유전자 상대밀도의 증가 값 $\Delta X_{LL}(t = 1)$ 이 네 상태 모델에서 가장 크며, 두 상태 모델에서 가장 작고 세 상태 모델에서 중간이기 때문이다.



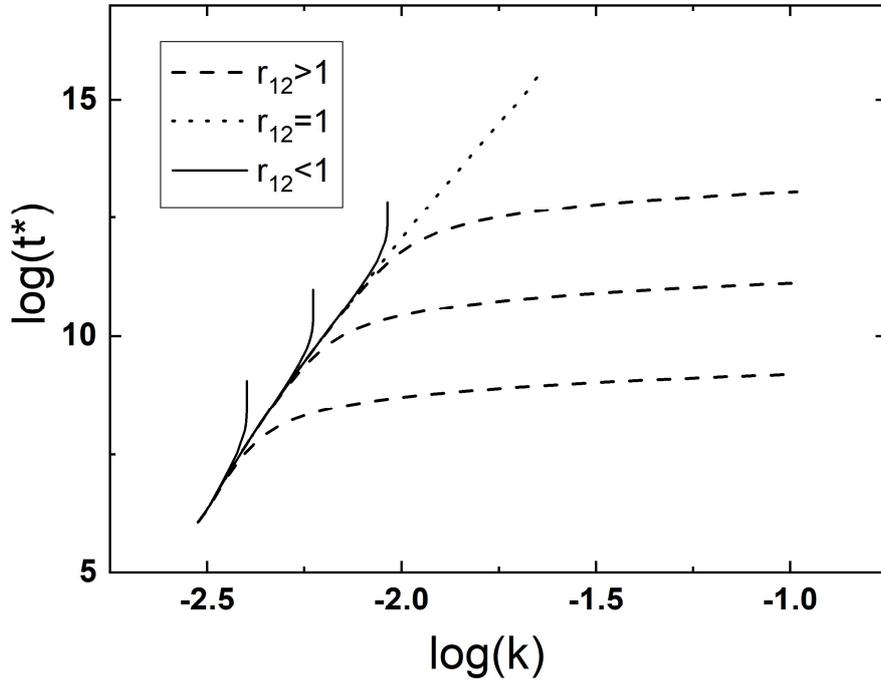
[그림 3] 두, 세, 네 상태 HCDMS 모델에서 $L = 10$, $\mu = 0.001$ 일 때 적응도땀음변수의 로그값 $\log(k)$ 에 따른 진화시간의 로그값 $\log(t^*)$ 변화. Jukes-Cantor 돌연변이 체계에서 $\log(t^*)$ 의 컴퓨터 시뮬레이션 결과는 밑에서 위로 선택이익, $s = 10^{-5}, 10^{-7}, 10^{-9}, 0$ 에 대한 것을 나타내며 유전자서열요소의 허용 상태 수가 두 개인 경우($n=2$)는 대시로, 세 개인 경우($n=3$)은 점선으로, 네 개인 경우($n=4$)는 실선으로 그려져 있다.

진화시간 t^* 는 $\Delta X_{LL}(t^*) = 0.01$ 이란 기준을 만족시키는 시간 간격으로 측정되었다. [그림 3]은 Jukes-Cantor 돌연변이 체계를 사용하는 두, 세, 네 상태 HCDMS 모델에서 $L = 10$, $\mu = 0.001$ 이고 $s \geq 0$ 인 다양한 선택이익에 대하여 적응도땀음변수의 로그값 $\log(k)$ 에 따른 진화시간의 로그값 $\log(t^*)$ 의 변화를 보여준다. [그림 3]에서 보이듯이 오차문턱은 두 상태 모델에서 $k \geq \mu$, 세 상태 모델에서 $k \geq 2\mu$, 네 상태 모델에서 $k \geq 3\mu$ 과 같은 제한을 부과한다. [그림 3]에서 $s > 0$ 에서 진화시간은 큰 적응도땀음변수에서 포화되었다. 이는 동일한 s 일 때 $\Delta X_{LL}(t)$ 는 동일한 특성 시간 $\tau (\approx 1/s)$ 후에 적응도땀음변수와 무관한 지수적 시간 의존성을 갖기 때문이다. $s = 0$ 일 때 진화시간은 적응도땀음변수의 맥 법칙에 따라 변화한다. 네 상태 모델에서 $\mu \ll k \ll 1$ 일 때 진화시간은 두, 세 상태 모델과 유사하다. [그림 3]은 네 상태 모델의 진화시간이 두, 세 상태 모델의 동일한 μ 및 k 일 때의 진화시간보다 짧음을 보여 주며, 이는 두, 세 상태 모델보다 네 상태 모델에서 개체들이 역전대립유전자에 더 빨리 도착한다는 것을 나타낸다.



[그림 4] 두, 세, 네 상태 HCDMS 모델에서 $L = 10$, $\mu = 0.001$ 일 때 적응도맷음변수의 로그값 $\log(k)$ 에 따른 진화시간의 로그값 $\log(t^*)$ 의 변화. Jukes-Cantor 돌연변이 체계에서 $\log(t^*)$ 의 컴퓨터 시뮬레이션 결과는 왼쪽에서 오른쪽으로 선택이익, $s = -10^{-7}$, -10^{-9} , -10^{-11} 에 대한 것을 나타내며 유전자서열요소의 허용 상태 수가 두 개인 경우($n = 2$)는 대시로, 세 개인 경우($n = 3$)는 점선으로, 네 개인 경우($n = 4$)는 실선으로 그려져 있다.

[그림 4]는 Jukes-Cantor 돌연변이 체계를 사용하고 두, 세, 네 상태 HCDMS 모델에서 $L = 10$, $\mu = 0.001$ 이고 다양한 음의 선택이익 $s < 0$ 에 대하여 적응도맷음변수의 로그값 $\log(k)$ 에 따른 진화시간의 로그값 $\log(t^*)$ 의 변화를 보여준다. [그림 4]는 네 상태 모델에서 $s < 0$ 일 때 진화시간은 [그림 3]에 나타난 것처럼 두, 세 상태 모델에서의 진화시간보다 짧았다. 적응도맷음변수가 임계적응도맷음변수 k_c 에 접근할 때 $s < 0$ 일 때의 진화시간은 발산한다. 임계적응도맷음변수 k_c 는 선택이익의 절댓값 $|s|$ 가 감소함에 따라 증가한다. [그림 4]에서 네 상태 모델의 임계적응도맷음변수는 두, 세 상태 모델의 임계적응도맷음변수보다 크다.

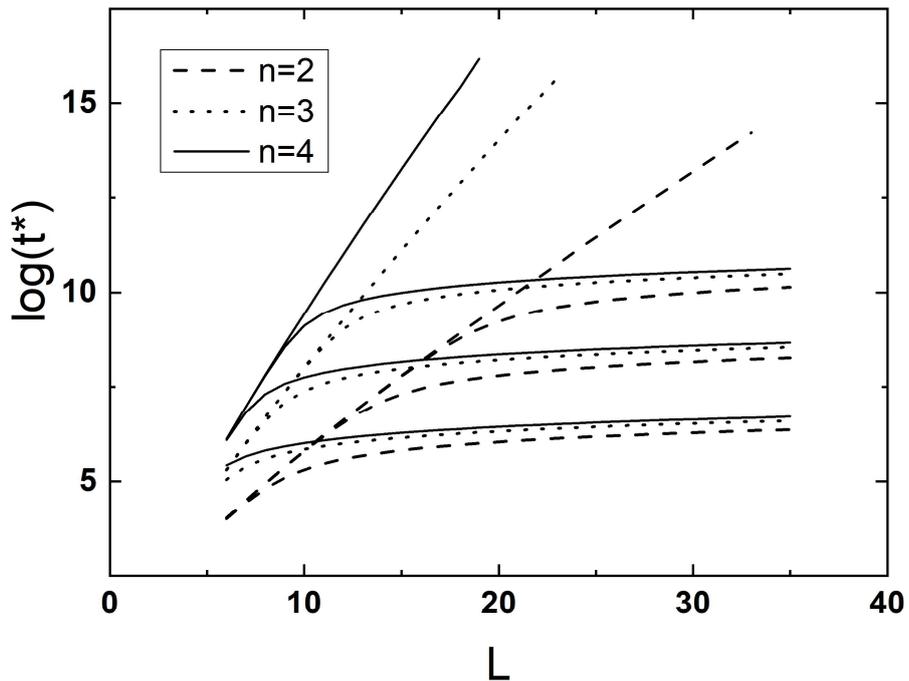


[그림 5] 네 상태 모델에서 $L = 10, s = 0, \mu = 0.001, r_{23} = r_{42} = 1$ 일 때 적응도맷음변수의 로그값 $\log(k)$ 에 따른 진화시간의 로그값 $\log(t^*)$ 변화. $\log(t^*)$ 의 컴퓨터 시뮬레이션 결과는 밑에서 위로 돌연변이율의 상대 비율 $r_{12} = 1 + 10^{-6}, 1 + 10^{-8}, 1 + 10^{-10}$ 에 대한 것을 나타내며 대신으로 그려져 있고, $r_{12} = 1$ 에 대해서는 점선으로 그려져 있고, 왼쪽에서 오른쪽으로 $r_{12} = 1 - 10^{-6}, 1 - 10^{-8}, 1 - 10^{-10}$ 에 대해서는 실선으로 그려져 있다.

[그림 5]는 네 상태 HCDMS 모델에서 $L = 10, s = 0, \mu = 0.001, r_{23} = r_{42} = 1$ 이고 다양한 r_{12} 에 대하여 적응도맷음변수의 로그값 $\log(k)$ 에 따른 진화시간의 로그값 $\log(t^*)$ 의 변화를 나타낸다. [그림 5]에서 $r_{12} > 1$ 인 경우에, 진화시간은 큰 적응도맷음변수에서 포화되었다. 이것은 [그림 3]에서 양의 선택이익 $s > 0$ 을 갖는 진화시간과 유사하다. 진화시간이 포화된 이유는 동일한 r_{12} 일 때 $\Delta X_{LL}(t)$ 가 동일한 특성시간 $\tau (\approx 1/|dr_{12}|)$ 후에 시간이 증가함에 따라 지수적 시간 의존성을 갖기 때문이다. 큰 적응도맷음변수에서 $r_{12} > 1$ 일 때 진화시간은 dr_{12} 가 1/100로 감소할 때 100배 증가한다. 이는 $r_{12} > 1$ 일 때 $\Delta X_{LL}(t)$ 이 특성시간 τ 까지 선형적 시간 의존성을 가지고, 이 특성시간의 크기는 돌연변이율의 비율의 절댓값 역수 $1/|dr_{12}|$ 에 비례하고, $\Delta X_{LL}(t)$ 는 특성시간 이후 시간이 증가함에 따라 기하급수적으로 증가하기 때문이다. $r_{12} < 1$ 인 경우에는 적응도맷음변수가 임계적응도맷음변수 k_c 에 가까워지면 진화시간이 발산되며, 임계적응도맷음변수는 돌연변이율의 비율의 절댓값이 감소함에 따라 증가한다. 추가적인 컴퓨터 시뮬레이션은 네 상태 모델의 진화시간이 r_{23} 과 관련이 없다는 것을 보여준다. 또한 추가적인 컴퓨터 시뮬레이션은

$dr_{42} = -dr_{12}$ 인 경우, $s = 0$, $r_{12} = 1$, $dr_{42} \neq 0$ 일 때의 진화시간이 $s = 0$, $r_{42} = 1$, $dr_{12} \neq 0$ 일 때의 진화시간과 거의 같음을 보여준다. 그러므로 네 상태 모델에서 진화시간은 돌연변이율의 비율 r_{12} 와 r_{42} 에 민감하게 의존하고, 진화시간에 대한 r_{12} 의 효과는 진화시간에 대한 r_{42} 의 효과와 반대이다.

초기 정상상태의 최적 대립유전자로부터의 평균 해밍거리로 정의된 \bar{d} 를 서열 길이 L 로 나눈 값으로 정의된 확장땀음변수 $E(=\bar{d}/L)$ 를 고정시키고, 유전자서열길이를 증가시키면서 계산하였다. 컴퓨터 시뮬레이션을 한 결과, 확장땀음변수를 고정시키고 유전자서열길이를 증가시키면 적응도땀음변수의 크기가 감소하고 오차문턱에 접근하였다. [그림 18]은 Jukes-Cantor 돌연변이 체계이고 두, 세, 네 상태 HCDMS 모델에서 $E = 0.1$, $\mu = 0.001$ 이고, $s \geq 0$ 인 다양한 선택이익에 대하여 유전자서열길이 L 에 따른 진화시간의 로그값 $\log(t^*)$ 의 변화를 보여준다. 고정된 확장땀음변수를 갖고 $s > 0$ 일 때 진화시간은 긴 서열 길이에서 포화된다. 이것은 동일한 s 일 때 동일한 특성 시간 $\tau (\approx 1/s)$ 후에 $\Delta X_{LL}(t)$ 가 지수적 시간 의존성을 가지고, 특성시간이 유전자서열 길이와 무관하기 때문이다.



[그림 6] 두, 세, 네 상태 HCDMS 모델에서 $E = 0.1$, $\mu = 0.001$, $s \geq 0$ 일 때 유전자서열길이 L 에 따른 진화시간의 로그값 $\log(t^*)$ 의 변화. Jukes-Cantor 돌연변이 체계에서 $\log(t^*)$ 의 컴퓨터 시뮬레이션 결과는 밑에서 위로 선택이익, $s = 10^{-5}$, 10^{-7} , 10^{-9} , 0에 대한 것을 나타내며 유전자서열요소의 허용 상태 수가 두 개인 경우($n=2$)는 대시로, 세 개인 경우($n=3$)은 점선으로, 네 개인 경우($n=4$)는 실선으로 그려져 있다.

[그림 6]에서 네 상태 모델의 진화시간은 두 상태 모델에 비해 더 길며 [그림 15]의 결과와는 반대이다. 이는 네 상태 모델의 대립유전자 분포가 두, 세 상태 모델의 대립유전자 분포보다 더 쉽게 비국소화되기 때문에, 같은 정도의 국소화된 분포를 얻기 위해서는 적응도뿔변수가 두, 세 상태 모델에 비해서 커야 한다. 따라서 상대적으로 큰 적응도뿔변수를 가진 네 상태 모델의 진화시간이 두, 세 상태 모델에 비해서 더 길어지게 된다. 네 상태 모델에서 $s=0$ 일 때 확장뿔변수를 고정시키고 유전자서열길이를 증가시키면 진화시간은 두, 세 상태 모델보다 더 큰 비율을 갖고 지수적으로 증가한다. 추가적인 컴퓨터 시뮬레이션은 $s < 0$ 일 때에, 네 상태 모델에서 확장뿔변수를 고정시키고 서열 길이를 증가시키면 유전자서열길이가 임계유전자서열길이보다 길 때에 진화시간이 관찰되지 않았다는 것을 제외하고는, $s=0$ 일 때의 진화시간과 거의 같음을 보여준다.

IV. 결론

진화이론자학자들이 진화를 설명하기 위하여 고안된 모델을 사용하여서 컴퓨터 시뮬레이션을 하였으며, 여러 진화의 작용력 중에서 돌연변이와 자연선택을 고려한 돌연변이-자연선택 모델을 사용하였고, 단순지형에서 단성생식을 하는 반수체를 고려하였다. 돌연변이로는 염기치환을 고려하였고, 생식과정에서의 복제오류에 의한 돌연변이를 고려한 결합모델을 살펴보았다. 자연 상태에서 발견되는 대립유전자인 야생형 S_0 에만 높은 적응도를 부여하고 나머지 대립유전자들에게는 낮은 적응도를 부여하는 단순지형에서 돌연변이율이 그 이상 커지면 대립유전자 분포가 무질서한 상태로 바뀌는 오차문턱의 존재가 확인되었다.

진화시간을 계산하기 위하여 단순지형에서 대립유전자 빈도분포가 야생형 근처에 국소화된 정상상태에 도달한 후에, 적응지형도를 갑자기 단순지형에서 대립유전자 S_0^* 에 큰 적응도를 부여하는 단순비대칭지형으로 바꾸었다. 야생형 근처에 국소화되었던 분포로부터 단순비대칭지형으로 바뀌고 난 후에 전체 개체수의 1%가 역전대립유전자 S_0^* 를 갖는데 걸리는 시간, 즉 진화시간을 DNA길이, 개체수, 비대칭변수 등을 변화시키며 계산하였다.

본 논문은 개체수가 무한한 반수체 결합 불연속-시간 돌연변이-자연선택(HCDMS) 모델에서 유전자서열요소의 허용 상태 개수에 따른 여러 가지 특성의 변화를 조사하였다. 이 모델은 무성 생식으로 번식하고 시간이 세대별로 불연속적이며 점 돌연변이만을 고려하였다. n -상태 HCDMS 모델의 단순지형에서 오차문턱은 $k \ll 1$ 일 때 $k/(n-1)$ 로 주어졌으며, k 는 적응도뿔변수이다. 단순지형의 정상상태에서 비대칭 단순지형으로 갑자기 바꾼 후에, 역전대립유전자로의 진화시간 및 상대밀도의 시간적 변화를 조사하였다.

네 상태 모델에서 두 유전자서열요소 사이의 돌연변이율이 같지 않은 경우도 조사하였다. 유전자서열요소 i, j 사이의 돌연변이율 μ_{ij} 에 대해서 $\mu_{ij} = \mu_{ji}$, $\mu_{12} = \mu_{13}$, $\mu_{42} = \mu_{43}$ 로 가정하였다. 돌연변이율 μ_{14} 는 두, 세 상태 모델에서 돌연변이율 μ 와 동일한 기준 값으로 설정되었고, 돌연변이율의 상대 비율 r_{ij} 는 μ_{ij}/μ_{14} 로 정의되었고, 돌연변이율의 상대 비율 차 dr_{ij} 는 $r_{ij} - 1$ 로 정의되었다.

모든 유전자서열요소 i 와 j 에 대한 돌연변이율이 같은 Jukes-Cantor 돌연변이체계인 경우에 네 상태 모델에서 역전대립유전자의 상대밀도 증가 $\Delta X_{LL}(t)$ 은 특성시간 $\tau (\approx 1/|s|)$ 까지 선형적 시간의존성을 갖고, 특성시간 후에는 $s > 0$ 일 때 지수적으로 증가하였고, $s < 0$ 일 때는 포화되었다. 여기서 s 는 선택이익이다. $r_{12} \neq 1$, $s = 0$, $r_{23} = r_{42} = 1$ 일 때의 $\Delta X_{LL}(t)$ 는 특성시간 $\tau (\propto 1/|dr_{12}|)$ 까지 선형적으로 시간에 따라 증가했고, 특성시간 후에는 $r_{12} > 1$ 일 때 지수적으로 증가하였고, $r_{12} < 1$ 일 때 포화되었다. 추가로 한 시뮬레이션에서는 $dr_{42} = -dr_{12}$ 라면, $r_{12} = 1$, $dr_{42} \neq 0$ 일 때의 $\Delta X_{LL}(t)$ 는 $r_{42} = 1$, $dr_{12} \neq 0$ 일 때의 $\Delta X_{LL}(t)$ 와 거의 같았다. 그러므로 네 상태 HCDMS 모델에서 상대밀도의 시간에 따른 변화는 돌연변이율의 상대 비율인 r_{12} 와 r_{42} 에 매우 민감하게 의존하였다.

네 상태 모델에서 Jukes-Cantor 돌연변이체계일 때에 적응도뿔변수를 증가시키면, 진화시간은 $s > 0$ 이면 큰 적응도뿔변수에서 포화되었고, $s = 0$ 이면 진화시간이 적응도뿔변수의 역

법칙(power law)에 따라 변화되었고, $s < 0$ 일 때 진화시간은 적응도맷음변수가 임계적응도맷음 변수에 가까워질 때 발산되었다. 선택이익의 절댓값이 감소할 때에 임계적응도맷음변수가 증가함을 볼 수 있었다. 네 상태 모델이고 $s = 0, r_{23} = r_{42} = 1$ 일 때 적응도맷음변수를 증가시키면, 진화시간은 $r_{12} > 1$ 일 때 큰 적응도맷음변수에서 포화되었고, $r_{12} < 1$ 일 때 적응도맷음변수가 임계적응도맷음변수에 가까워질 때 진화시간은 발산된다. 뿐만 아니라 $dr_{42} = -dr_{12}$ 이라면 $r_{12} = 1, dr_{42} \neq 0$ 일 때의 진화시간은 $r_{42} = 1, dr \neq 0$ 일 때의 진화시간과 거의 같다. 네 상태 HCDMS 모델에서 진화시간은 돌연변이율의 비율 r_{12} 와 r_{42} 에 따라 매우 민감하게 변화함을 볼 수 있었다.

네 상태 모델에서 Jukes-Cantor 돌연변이체계를 갖고 확장맷음변수를 고정시키고 유전자서열길이를 증가시키면, $s > 0$ 에서 진화시간은 긴 유전자서열길이에서 포화되었고, $s = 0$ 일 때 진화시간은 유전자서열길이가 증가함에 따라 지수적으로 증가했고, $s < 0$ 에서는 유전자서열길이가 임계유전자서열길이보다 길 때에 진화시간이 관찰되지 않았다는 것을 제외하고는, $s = 0$ 일 때의 진화시간과 거의 같았다. 선택이익의 절댓값이 감소함에 따라 임계유전자서열길이는 증가하였다. 네 상태 모델에서 $s = 0, r_{23} = r_{42} = 1$ 인 경우 확장맷음변수를 고정시키고 유전자서열길이를 증가시키면, $r_{12} > 1$ 일 때 진화시간은 긴 유전자서열길이일 때 포화되었으며, $r_{12} < 1$ 일 때는 유전자서열길이가 임계유전자서열길이보다 길 때 진화시간이 관찰되지 않았다는 것을 제외하고는 $r_{12} = 1$ 에서의 진화시간과 거의 같았다. 돌연변이율의 비율 차의 절댓값 $|dr_{12}|$ 이 감소함에 따라 임계유전자서열길이는 증가했다. 만약 $dr_{42} = -dr_{12}$ 라면, $s = 0, dr_{12} > 0, r_{42} = 1$ 일 때의 진화시간은 $s = 0, dr_{42} < 0, r_{12} = 1$ 일 때의 진화시간과 거의 같았다. 네 상태 HCDMS 모델에서 상대 밀도의 시간적 변화와 진화시간과 같은 다양한 특성은 돌연변이율의 비율 차에 따라 민감하게 변화하였으며, 이는 선택이익과 유사한 효과를 보여준다. 돌연변이율의 비율 차에 대한 다양한 특성의 민감한 의존성은 이전의 논문에서 찾을 수 없던 새로운 이해를 준다.

본 연구에서 확장맷음변수를 고정시키고 유전자서열길이를 증가시킬 때에, 진화시간이 일정한 값으로 수렴하는 것을 볼 수 있었는데, 이러한 일정한 값으로의 수렴은 개체수가 무한대이기 때문에 나타나는 현상이라고 볼 수 있다. 두 상태 돌연변이-자연선택 모델에서 개체 수가 유한한 경우에 확장맷음변수를 고정시키고 유전자서열길이를 증가시키면 진화시간이 일정한 값으로 수렴하지 않고 지수적으로 증가하는 것을 볼 수 있었으며, 지수적으로 증가하는 비율은 선택이익 $s = 0$ 인 경우와 비슷했다. 개체수가 유한한 경우에 유전자서열길이를 어느 정도 이상 증가시키면 진화시간이 지수적으로 증가하는 현상은 두 상태 돌연변이-자연선택 모델에서 결합 모델과 분리 모델 모두에서 나타났기에, 네 상태 모델에서도 나타날 것으로 유추된다. 유전자서열길이가 증가함에 따라 진화시간이 지수적으로 증가하면, 아무리 긴 시간이 주어진다 하더라도 바뀌게 되는 유전자서열요소의 개수는 제한적이다. 예로서, 이전 논문에서 언급했듯이 대장균의 경우에 한 세대가 약 20분이기에 1억년 동안에 바뀔 수 있는 뉴클레오티드의 최대 개수는 개체수가 $N = 10^{10}$ 일 때에 대략 두 상태 결합 모델에서 46개, 분리 모델에서 49개에 불과했다. 따라서 이러한 최대 개수를 초과하여 뉴클레오티드가 바뀌어야 하는 새로운 유전자로의 진화는 불가능하다고 볼 수 있다. 또한 본 연구의 그림에서 밝힌 것처럼, 확장맷음변수를 고정시키고 유전자서열길이를 증가시키면, 진화시간의 지수적 증가 비율이 네 상태 모델에서 가장 컸으며, 두 상태 모델에서 가장

작았다. 이러한 결과는 유전자서열요소의 허용 상태 수가 실제처럼 네 개를 갖게 되면 진화시간이 더 빠르게 지수적으로 증가하기에, 진화가 되는 것이 더 어려워짐을 나타낸다. 결론적으로 말하면, 진화론자들이 고려하는 돌연변이, 자연선택 등의 어떠한 기제를 사용하더라도 원시생명체에서 인간이 되기까지 새로 만들어져야 할 수많은 유전자들의 출현을 도저히 설명할 수 없기에, 진화론을 설득력 있는 과학 이론으로 보기 어렵다.

이러한 결과를 통해서 진화론이 도킨스와 같이 설정 자체를 잘못하여 생기는 오류뿐만 아니라 이 설정을 바로잡아 실험하여도 심각한 오류가 있음을 알 수 있다. 그럼에도 불구하고 세상에서는 진화론이 논리적으로 충분한 타당성을 갖춘 것처럼 포장되어지고, 거기에 대해서 의문을 품으면 과학적 소양이 부족한 것처럼 매도함으로써, 진화론에 대한 충분한 이해와 연구 없이 맹목적으로 믿게 만드는 또 하나의 사이비 종교와 비슷하게 되었다. 이런 세대 가운데 본 연구 결과가 하나님의 창조섭리를 밝히는 작은 빛이 되길 바란다.

참고문헌

- [1] C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski and C. Adaml, *Nature* 412, 331 (2001).
- [2] Y. S. Semenov and A. S. Novozhilov, *Math. Biosci.* 266, 1 (2015).
- [3] L. Yuan, *Math. Biosci.* 285, 61 (2017).
- [4] M. Eigen, *Naturwissenschaften* 58, 465 (1971).
- [5] M. Eigen, J. McCaskill and P. Schuster, *J. Phys. Chem.* 92, 6881 (1988).
- [6] I. Leuthäuser, *J. Stat. Phys.* 48, 343 (1987).
- [7] P. Tarazona, *Phys. Rev. A* 45, 6038 (1992).
- [8] S. Franz and L. Peliti, *J. Phys. A: Math. Gen.* 30, 4481 (1997).
- [9] B. Drossel, *Adv. Phys.* 50, 209 (2001).
- [10] K. S. Lee and W. Gill, *Int. J. Mod. Phys. C* 18, 1985 (2007).
- [11] M. Eigen, J. McCaskill and P. Schuster, *J. Phys. Chem.* 92, 6881 (1988).
- [12] M. Eigen, *Naturwissenschaften* 58, 465 (1971).
- [13] M. Eigen, J. McCaskill, and P. Schuster, *J. Phys. Chem.* 92, 6881 (1988).
- [14] I. Leuthäusser, *J. Stat. Phys.* 48, 343 (1987).
- [15] I. Leuthäusser, *J. Chem. Phys.* 84, 1884 (1986).
- [16] P. Tarazona, *Phys. Rev. A* 45, 6038 (1992).
- [17] P. G. Higgs, *Genet. Res. Camb.* 63, 63 (1994).
- [18] S. Franz and L. Peliti, *J. Phys. A: Math. Gen.* 30, 4481 (1997).
- [19] T. Wiehe, *Genet. Res. Camb.* 69, 127 (1997).
- [20] S. Galluccio, *Phys. Rev. E* 56, 4526 (1997).
- [21] R. Bürger, *Genetica* 102/103, 279 (1998).
- [22] B. Drossel, *Adv. Phys.* 50, 209 (2001).
- [23] E. Baake, *J. Biol. Syst.* 3, 343 (1995).
- [24] E. Baake, M. Baake and H. Wagner, *Phys. Rev. Lett.* 78, 559 (1997).
- [25] H. Wagner, E. Baake and T. Gerisch, *J. Stat. Phys.* 92, 1017 (1998).