

진화모델에 대한 최근의 시뮬레이션 결과 요약 및 고찰*

길원평(부산대학교 물리학과 교수)

논문초록

진화이론학자들이 진화를 설명하기 위하여 고안된 모델을 사용하여서 컴퓨터 시뮬레이션으로 진화시간을 계산한 최근의 연구 결과들을 요약하고 고찰해 보았다. 여러 진화의 작용력 중에서 돌연변이와 자연선택을 고려한 돌연변이-자연선택 모델에서 단성생식을 하는 반수체를 고려하였으며, 돌연변이로는 염기치환을 고려하였다. 생식과정에서의 복제오류에 의한 돌연변이를 고려한 결합모델과 돌연변이원에 의해서 일생 중 어느 때라도 가능한 돌연변이를 고려한 분리모델 모두를 살펴보았다. 단순지형에서 대립유전자 빈도분포가 야생형 근처에 국소화된 정상상태에 도달한 후에, 적응지형을 단순지형에서 야생형의 모든 뉴클레오티드가 치환된 대립유전자에 야생형보다 큰 적응도를 부여하는 단순비대칭지형으로 바꾸고, 전체 개체수의 1%가 야생형의 모든 뉴클레오티드가 치환된 대립유전자를 갖는데 걸리는 시간, 즉 진화시간을 DNA길이, 개체수, 비대칭변수 등을 변화시키며 계산하였다. 개체수가 유한하고 비대칭변수 $r > 0$ 일 때에 DNA길이가 임계DNA길이 이상 길어지면 DNA길이가 길어짐에 따라 진화시간이 지수적으로 증가하였으며, 지수적인 증가율은 개체수와 비대칭변수에 무관하였다. 또한 비대칭변수가 100 이상 커지면 비대칭변수가 더 증가하더라도 진화시간은 더 이상 감소하지 않았다. 요약된 연구 결과의 핵심은 새로운 유전정보를 가진 유전자를 만들기 위하여 바뀌어져야 할 뉴클레오티드의 개수가 증가함에 따라 진화시간이 지수적으로 증가하기에, 새로운 유전자가 아무리 큰 적응도를 가진다 하더라도 기존의 유전자로부터 60개보다 많은 뉴클레오티드가 바뀌어야만 한다면, 그 새로운 유전자로의 진화는 불가능하다는 것이다. 진화론자들이 돌연변이, 자연선택 등을 사용하여서 원시생명체에서 인간이 생겨났다는 생물학적 진화론을 설명하고자 노력하지만, 받아들이지 않으면 안 될 정도로 설득력있는 과학이론은 아니다.

주제어: 진화시간, 돌연변이-자연선택 모델, 시뮬레이션, 단순지형, 생물학적 진화론

* 이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었습니다. 이러한 연구를 하도록 시작하게 하시고 영감을 주셔서 결과를 얻게 하신 하나님께 감사드리며, 논문을 읽고 유익한 조언을 해주신 전태수 교수께 감사드립니다.

* 2011년 3월 11일 접수, 6월 15일 수정, 6월 21일 게재확정.

I. 서론

기독교 신앙은 하나님의 창조를 기초로 하고 있다. 성경의 창세기 첫 구절이 기술하듯이 전지전능하신 하나님께서 천지 만물을 창조하시고, 특히 인간을 하나님의 형상대로 존귀하게 창조하신 것을 기독교 신앙의 기본으로 삼고 있다. 그런데, 우연히 결합된 원시생명체로부터 인간이 생겨났다는 진화론이 전 세계의 교육현장에서 가르쳐지고 있다. 이로 인하여 현대의 많은 기독교인들은 창조에 대한 확신을 갖지 못하고, 많은 신학자들은 성경을 정확무오한 하나님의 말씀으로 받아들이지 못하고 있다. 믿음의 선조국가이었던 유럽과 미국에서는 기독교의 순수하고 철저한 신앙이 무너지고 있으며, 현대의 기독교는 큰 위기를 맞이하고 있다. 이러한 위기에 대한 여러 다른 이유들도 있겠지만, 필자의 생각에는 진화론을 신봉하는 무신론자들, 예를 들어서 ‘만들어진 신’이란 책을 쓴 리처드 도킨스와 같은 학자들에 의해서 기독교는 강력한 도전에 직면하고 있으며(Dawkins, 2007), 기독교 진영은 그것을 제대로 대응하지 못하고 있기 때문이라고 본다. 또한 일반인들은 무신론자들의 말을 더 합리적이며 과학적이라고 인식하고 설득을 당하며 기독교로부터 멀어지고 있다.

물론 기독교는 예수님의 동정녀 탄생과 부활과 같은 과학으로는 도저히 설명할 수 없는 초과학적인 신비를 믿는 종교이지만, 그래도 이미 믿음을 가진 기독교인과 불신자를 위하여 어느 정도의 합리적인 논리도 제공해 주어야 한다. 왜냐하면 모든 인간은 이성을 갖고 진리를 따르길 원하기에, 기독교신앙이 분명한 진리에 근거하고 있다는 안도감을 갖도록 도움을 주어야 하며, 하나님의 진리는 모든 학문 안에서도 마땅히 진리로서 드러나야 하기 때문이다. 이러한 이유에서 학문 안에서 하나님의 진리를 진리로서 드러나도록 변증해야 하는 크리스천 학자의 사명은 매우 크다.

본 논문은 진화모델에 대한 최근의 컴퓨터 시뮬레이션 결과들을 소개함으로써(Gill, 2010a; 2010b), 진화론이 받아들이지 않으면 안 될 정도로 설득력있는 과학이론은 아니라는 사실을 보여 주고자 한다. 진화이론학자들이 진화를 설명하기 위하여 고안한 모델을 그대로 사용하여 진화를 하는데 걸리는 시간을 계산함으로써 진화가 확률적으로 불가능하다는 것을 보이려고 한다. 진화론 안에도 여러 영역이 있기에, 본 논문은 원시생명체에서 인간으로 진화하였다고 보는 생물학적 진화론에 국한하고자 한다. 이러한 진화가 이루어지려면, 진화론에서 제시하고 있는 여러 진화의 작용력들인, 돌연

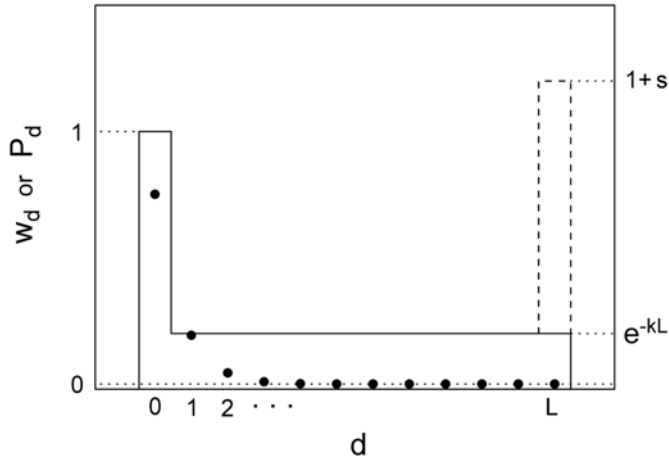
변이, 자연선택, 유전적부동(genetic drift) 등의 과정을 통하여서 원시 생명체의 간단한 유전체(genome)로부터 복잡하고 다양한 기능을 갖춘 유전체로 바뀌어져야 한다. 최근 들어서 컴퓨터 계산속도가 빨라짐으로서, 진화이론학자들이 고안한 모델을 사용하여서 진화시간을 계산하는 것이 가능하게 되었다. 따라서 자연계에는 전혀 존재하지 않았던, 새로운 기능을 가진 유전체로 변화하는데 걸리는 시간, 즉 진화시간에 대한 최근의 컴퓨터 시뮬레이션 결과들을 소개하고자 한다.

II. 모델 및 시뮬레이션 결과

본 논문에서는 여러 진화의 작용력 중에서 돌연변이와 자연선택을 고려하는 돌연변이-자연선택 모델(mutation-selection model)에 대한 연구결과를 소개하고자 한다(Hofbauer, 1985; Bürger, 1998; Gill, 2010a; 2010b). 단성생식(asexual reproduction)을 하는 반수체(haploid)를 고려하였고, 집단(population)의 개체수 N 은 모든 세대에 걸쳐서 일정하게 유지된다고 가정하였다. 우리가 계산한 DNA길이 L 은 유전체의 전체 길이에 비해서는 너무 짧기에, 특정 유전자(gene)의 DNA 일부분이 바뀌어져서 새로운 기능을 갖는 유전자로 된다고 가정할 때에, 길이 L 인 DNA는 새로운 기능을 갖는 유전자가 되기 위하여 바뀌어져야 할 DNA 부분만을 지칭한다고 보았다. 참고로 대장균의 유전체는 약 464만 염기쌍으로 이루어져 있고, 평균 유전자 크기는 1000 염기쌍이다(Pierce, 2009: 211). 실제 DNA는 네 종류의 뉴클레오티드로 구성되지만, 우리는 간단히 두 종류, 즉 +1과 -1로 구성된다고 가정하였다.

DNA서열은 돌연변이에 의해 바뀌어진다고 가정하였다. 돌연변이는 생식과정에서의 복제오류에 의해서 생겨날 수도 있고, 자외선, 화학물질 등의 돌연변이원(mutagen)에 의해서 생겨날 수 있다. 생식과정에서의 돌연변이를 고려하는 모델을 결합 돌연변이-자연선택 모델(coupled mutation-selection model)이라고 부르며(Higgs, 1994; Franz and Peliti, 1997; Burger, 1998; Drossel, 2001), 돌연변이원에 의해서 일생 중 어느 때라도 가능한 돌연변이를 고려하는 모델을 분리 돌연변이-자연선택 모델(decoupled mutation-selection model)이라고 부른다(Baake, 1997; Wagner, 1998). 본 논문에서는 간략히 결합모델과 분리모델이라고 부르겠다. 돌연변이로는 DNA의 뉴클레오티드 하

나가 바뀌는 염기치환(base substitution)을 고려하였고, DNA의 모든 뉴클레오티드는 동일한 돌연변이율 μ 를 갖는다고 가정하였다.

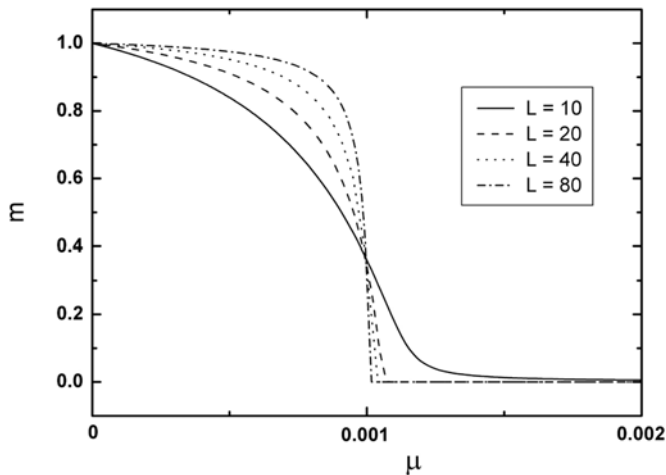


[그림1] 단순지형에서 해밍거리(d)에 따른 대립유전자의 적응도(W_d) 및 빈도(P_d).

자연선택을 정량화하기 위하여 대립유전자(allele)의 상대적 번식능력을 나타내는 적응도(fitness)를 도입하고, 집단 내의 모든 가능한 대립유전자에 대한 적응도를 나타낸 것을 적응지형도(fitness landscape)라고 부른다. 자연상태에서 발견되는 대립유전자인 야생형(wild type) $S_0=(+1, +1, \dots, +1)$ 에만 높은 적응도 $W_0(=1)$ 를 부여하고, 나머지 대립유전자들에게는 낮은 적응도 $W_1(=e^{-kL})$ 을 부여하는 단순지형(simple landscape, or, sharply-peaked landscape)을 고려하였다. 여기서 k 를 적응도변수(fitness parameter)라고 부른다. 단순지형은 야생형의 뉴클레오티드서열 중에서 하나라도 돌연변이를 일으키면 자연에 적응하는 좋은 기능을 상실하여 적응도가 감소하는 경우에 대응이 되기, 우리가 고려하는 DNA부위는 아미노산 서열을 암호화하는 엑손(exon)에 해당하는 것이라고 볼 수 있다. 단순지형은 [그림1]에서 실선으로 표현되었는데, x축은 야생형 S_0 를 기준하였을 때 대립유전자가 갖는 돌연변이 개수 d 를 나타내고, y축은 그 대립유전자의 적응도 W_d 를 나타낸다. 돌연변이 개수 d 를 야생형으로부터의 해밍거리(Hamming distance)라고 부른다. 단순지형을 사용하여 여러 세대동안 시뮬레이션을 하면, 집단의 대립유전자 빈도분포는 야생형 근처에 국소화된 정상상태(localized

stationary state)에 도달한다.

[그림1]에서 둥근 부호들은 국소화된 정상상태의 대립유전자 빈도분포의 한 예를 표현하는데, x축은 여전히 야생형으로부터의 해밍거리 d 를 뜻하고, y축은 해밍거리 d 를 갖는 모든 대립유전자들의 총 빈도 P_d 를 나타낸다. 이렇게 야생형 근처에 국소화된 분포를 유사종(quasispecies)이라고 부른다(Eigen et al., 1988). 하지만 돌연변이율을 증가시켜서 임계값보다 크게 만들면, 대립유전자 빈도분포는 더 이상 야생형 근처에 국소화되지 않고 무질서한 상태로 바뀌게 된다. 즉, DNA가 갖고 있었던 유전정보를 상실하게 되며, 이 임계값 μ_c 를 오차문턱(error threshold)이라고 부른다.



[그림2] 단순지형이며 여러 DNA길이(L)에서 돌연변이율(μ)에 따른 질서맞음변수(m).

[그림2]의 x축은 돌연변이율 μ 를 나타내고, y축은 질서맞음변수(order parameter)

$m=(1-2\langle d \rangle/L)$ 을 나타내며, $\langle d \rangle (= \sum_{d=0}^L P_d \cdot d)$ 는 평균해밍거리를 뜻한다. 만약 집단

내의 모든 대립유전자가 야생형 S_0 이라면 평균해밍거리 $\langle d \rangle=0$ 이고 $m=1$ 이 되며, 집단 내의 대립유전자 분포가 무질서한 상태가 되면 평균해밍거리 $\langle d \rangle=L/2$ 이고 $m=0$ 이 된다. [그림2]는 개체수가 무한한 경우를 나타내며, 돌연변이율이 어떤 특정한 값보다 커질 때에 질서맞음변수 m 이 급격히 0으로 떨어지는 것을 볼 수 있다. DNA길이 L 이 길어질수록 더욱 급격하게 변하는 것을 볼 수 있으며, L 이 무한대가 되면 임계값에서

일차상전이(first-order phase transition)를 하는 것으로 알려져 있다(Franz and Peliti, 1997). 단순지형에서 임계값 μ_c 에 대한 이론값은 다음과 같다.

$$\mu_c \approx \frac{1}{L} \ln\left(\frac{W_0}{W_1}\right) = \frac{1}{L} \ln(e^{kL}) = k$$

본 논문에서는 $W_0/W_1=e^{kL}$ 이라고 가정하였기에, 임계값 μ_c 는 DNA길이 L 에 무관하게 적응도변수 k 와 같다. 이렇게 임계값 μ_c 를 DNA길이 L 에 무관하게 만들기 위하여, 앞에서 언급하였듯이 $W_0=1$ 과 $W_1=e^{-kL}$ 이라고 놓았다. [그림2]로부터 $\mu_c=k(=0.001)$ 임을 쉽게 볼 수 있다. 만약 W_0/W_1 의 값을 고정시키고 DNA길이를 증가시키면 임계값 μ_c 가 감소하여 실제 생명체들의 돌연변이율보다도 작아지기 때문에, DNA길이를 어느 정도 이상으로는 증가시킬 수 없게 된다. 단순지형에서는 결합모델과 분리모델 모두에서 야생형 근처에 국소화된 유사종 형성과 오차문턱의 존재가 확인되었다(Eigen et al., 1988; Baake and Gabriel, 2000).

진화시간을 계산하기 위하여 단순지형에서 대립유전자 빈도분포가 야생형 근처에 국소화된 정상상태에 도달한 후에, 적응지형도를 갑자기 단순지형에서 단순비대칭지형으로 바꾸었다. 단순비대칭지형은 $d \neq L$ 일 때의 적응도는 단순지형과 같으며, $d=L$ 일 때의 적응도 $W_L(=1+s)$ 는 야생형보다 높다고 가정한다. 여기서 s 를 선택계수(selection coefficient)라고 부르며, 본 논문에서는 $s=r(1-e^{-kL})$ 로 놓고, r 을 비대칭변수(asymmetric parameter)라고 하였다. $r=0$ 이면 적응지형도가 대칭이며($W_0=W_L$), $r \neq 0$ 이면 비대칭이다($W_0 \neq W_L$). 단순비대칭지형의 $d=L$ 일 때의 적응도를 [그림1]에서 대시선으로 나타내었다. [그림1]에서 보듯이 단순비대칭지형에서는 두 개의 적응정점(fitness peak)이 낮은 적응도를 가지는 적응계곡(fitness valley)에 의해서 분리되어 있다.

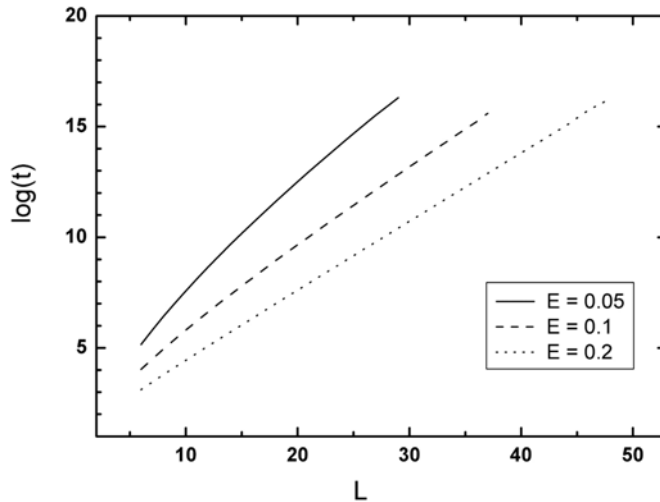
$d=L$ 인 경우는 야생형의 모든 뉴클레오티드가 치환된 대립유전자 $S^*_0=(-1, -1, \dots, -1)$ 에 대응되므로, 단순비대칭지형은 야생형의 모든 뉴클레오티드가 돌연변이에 의해 바뀌어졌을 때에 자연에 더 적응하는 새로운 기능을 가져서 높은 적응도를 갖는다고 가정한다. 따라서 야생형 S_0 근처에 국소화되었던 분포가 적응계곡을 통과하여서 S^*_0 로 옮겨 가는 것을 자연에 적응하는 새로운 기능을 갖게 되는 진화로 해석할 수 있다. 여기서 상기시키고 싶은 점은, 우리가 고려하는 길이 L 인 DNA는 야생형으로부터 새로

운 기능을 가져서 높은 적응도를 가진 유전자가 되기 위하여 바뀌어져야 할 DNA 부분만을 지칭한다는 것이다.

우리는 적응지형도가 단순비대칭지형으로 바뀌고 난 후부터, 전체 개체수의 1%가 대립유전자 S_0^* 를 갖는데 걸리는 시간 간격, 즉 진화시간 t 를 계산하였다. 초기 대립유전자 분포의 야생형 근처에 국소화된 정도를 일정하게 유지하기 위하여, 퍼짐변수 (extension parameter) $E(= \langle d \rangle / L)$ 값을 고정시키고 DNA길이 L 을 변화시키면서 진화시간 t 를 계산하였다. 참고로 퍼짐변수를 계산할 때에는 단순지형에서 구한 평균해밍거리 $\langle d \rangle$ 를 사용하였다. 즉, 여러 DNA 길이에 대하여 단순지형에서 같은 정도로 국소화된 대립유전자 분포를 초기분포로 놓고, 단순비대칭지형을 도입한 후에 진화되는데 걸리는 시간을 계산하였다.

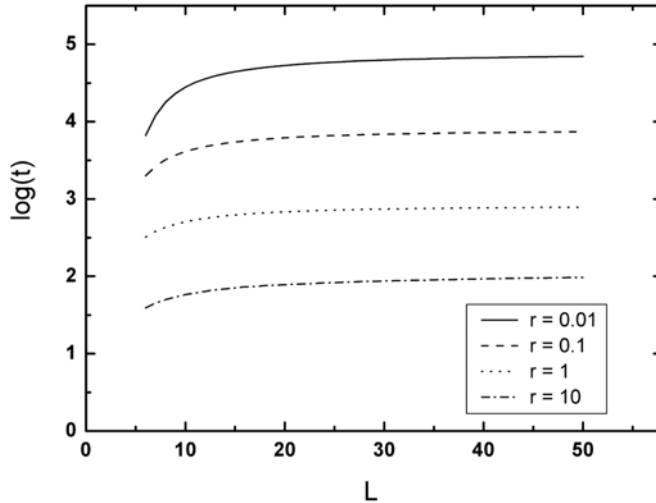
이제부터는 최근에 길원평에 의하여 컴퓨터 시뮬레이션으로 진화시간을 계산한 결과들을 요약하여 소개하고자 하며(Gill, 2010a; 2010b), 요약된 결과들에 대한 이해를 돕기 위하여 학술지에 발표했던 그림들과는 다르게 재구성하여 아래의 그림들을 만들었다. [그림3]은 결합모델을 사용하여서 개체수가 무한하며 비대칭변수 $r=0$ 일 때의, 즉 야생형 S_0 의 적응도 W_0 와 대립유전자 S_0^* 의 적응도 W_L 이 같을 때의, 진화시간 t 를 나타낸다. [그림3]의 x축은 DNA길이 L 을 나타내고, y축은 진화시간 t 의 로그값, $\log(t)$ 을 나타내며, 돌연변이율 $\mu=0.001$ 로 놓았다. 여기서 단위 시간은 한 세대를 뜻한다. [그림3]에서 DNA길이 L 이 어느 정도 길어진 후에는 모든 그래프가 대략 직선임을 볼 수 있다. y축이 진화시간 t 의 로그값이므로 DNA길이 L 이 길어짐에 따라, 즉 새로운 기능을 갖기 위하여 바뀌어야 할 뉴클레오티드의 개수가 증가함에 따라, 진화시간이 지수적으로 증가한다는 것을 알 수 있다.

[그림3]에서 퍼짐변수 E 가 증가할수록 그래프의 기울기가, 즉 진화시간의 지수적인 증가율이, 감소하는 것을 볼 수 있다. 이 결과로부터 집단 내 대립유전자 빈도분포가 많이 퍼져 있을수록, 즉 야생형으로부터 돌연변이를 일으킨 대립유전자들의 빈도가 증가할수록, 더 쉽게 진화가 이루어진다는 것을 알 수 있다.



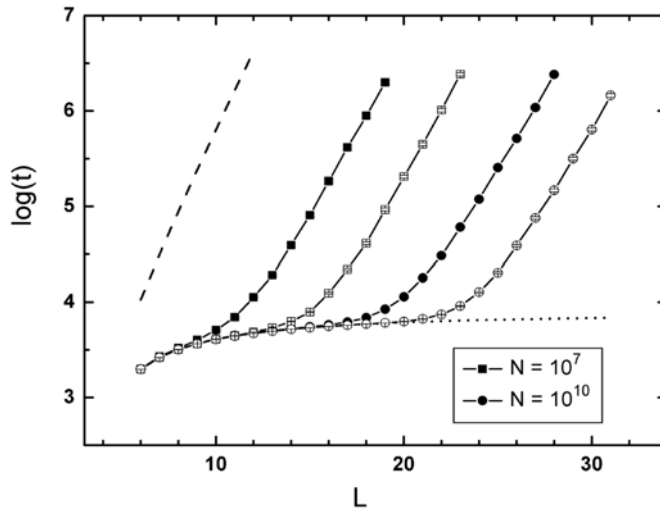
[그림3] 결합모델에서 개체수가 무한하며 $r=0$ 일 때 DNA길이(L)에 따른 진화시간(t).

[그림4]는 결합모델을 사용하여서 개체수가 무한하며 비대칭변수 $r>0$ 일 때의, 즉 $W_L > W_0$ 일 때의, 진화시간 t 를 나타내며, 돌연변이율 $\mu=0.001$, 퍼짐변수 $E=0.1$ 로 놓았다. [그림4]에서 비대칭변수 $r>0$ 일 때에 DNA길이 L 이 어느 정도 이상 길어지면, DNA 길이가 더 길어지더라도 진화시간은 더 이상 증가하지 않고 거의 일정한 값을 갖는 것을 볼 수 있다. 즉, DNA 길이가 아무리 길어지더라도 비슷한 시간 안에 진화가 이루어진다는 뜻이다. 하지만, 이러한 진화시간의 일정한 값으로의 수렴은 [그림5]에서 밝혀지듯이 개체수가 무한할 때에 나타나는 현상이다. [그림4]에서 비대칭변수 r 이 증가할수록 진화시간이 감소하는 것을 볼 수 있다. 야생형의 모든 뉴클레오티드가 치환된 대립유전자 S_0^* 의 적응도 W_L 이 야생형의 적응도 W_0 에 비하여 상대적으로 높은 적응도를 가지면 가질수록, S_0^* 로의 진화가 더 쉽게 이루어진다는 것을 뜻한다. 개체수가 무한할 때에 비대칭변수가 너무 크지 않으면, 즉 $r \ll 1$ 이면, 결합모델과 분리모델의 진화시간은 서로 매우 비슷한 값을 갖는 것으로 밝혀졌다(Gill, 2010b: 293; Baake and Gabriel, 2000).



[그림4] 결합모델에서 개체수가 무한하며 $r > 0$ 일 때 DNA길이(L)에 따른 진화시간(t).

[그림5]는 단순비대칭지형에서의 진화시간 t 를 나타내며, 돌연변이율 $\mu=0.001$, 퍼짐변수 $E=0.1$, 비대칭변수 $r=0.1$ 로 놓았다. [그림5]에서 점선은 개체수가 무한할 때의 진화시간을 나타내며, 결합모델과 분리모델의 결과가 너무 같아서 겹쳐 있다. [그림5]에서 부호가 있는 실선들은 개체수가 유한할 때의 진화시간을 나타내며, 속이 짝 찬 부호들은 결합모델의 결과를 나타내고, 속이 빈 부호들은 분리모델의 결과를 나타낸다. 네모 부호는 개체수 $N=10^7$ 을 나타내고, 둥근 부호는 $N=10^{10}$ 을 나타낸다. 개체수가 유한할 때에 시뮬레이션을 1000번 실행하여서 평균과 표준오차를 계산하였다. [그림5]에서 DNA길이 L 이 임계DNA길이(critical DNA length) 이상 길어지면 개체수가 유한할 때의 진화시간은 개체수가 무한할 때의 진화시간으로부터 벗어나서, DNA길이가 더 길어짐에 따라 지수적으로 증가하는 것을 볼 수 있다. [그림5]에서 결합모델에서는 $N=10^7$ 일 때에 대략 임계DNA길이 $L_c=8$ 이고, $N=10^{10}$ 일 때에 $L_c=16$ 이며, 분리모델에서는 $N=10^7$ 일 때에 대략 임계DNA길이 $L_c=12$ 이고, $N=10^{10}$ 일 때에 $L_c=20$ 임을 볼 수 있다. 그리고 개체수가 줄어들수록 더 짧은 임계DNA길이 L_c 에서, 개체수가 유한할 때의 진화시간이 개체수가 무한할 때의 진화시간으로부터 벗어나서 지수적으로 증가하기 시작하는 것을 볼 수 있다.



[그림5] 결합모델과 분리모델에서 개체수가 유한하며 $r > 0$ 일 때의 진화시간(t).
 속이 짝 찬 부호는 결합모델을, 속이 빈 부호는 분리모델을 나타낸다.
 네모 부호는 개체수 $N=10^7$ 을, 둥근 부호는 $N=10^{10}$ 을 나타낸다.

[그림5]에서 대시선은 개체수가 무한하며 비대칭변수 $r=0$ 일 때의 진화시간을 나타낸다. [그림5]에서 개체수가 유한할 때의 진화시간이 개체수가 무한할 때의 진화시간으로부터 벗어나서 어느 정도 증가한 후에는, 결합모델과 분리모델 모두에서 개체수가 유한할 때의 그래프들이 개체수와 무관하게 대시선과 비슷한 기울기를 가지는 것을 볼 수 있다. 따라서 개체수가 무한하며 비대칭변수 $r=0$ 일 때와 같은 비율로서 지수적으로 증가한다는 것을 알 수 있다. [그림5]로부터 $t=10^6$ 세대 안에 바뀔 수 있는 뉴클레오티드 최대 개수(L)는, 결합모델에서는 개체수 $N=10^7$ 이면 18개이며 $N=10^{10}$ 이면 27개이고, 분리모델에서는 $N=10^7$ 이면 21개이며 $N=10^{10}$ 이면 30개인 것을 볼 수 있다. 따라서 결합모델과 분리모델 모두에서 개체수가 1000배 증가하면 $t=10^6$ 세대 안에 바뀔 수 있는 뉴클레오티드의 최대 개수(L)가 대략 9개 증가한다. 앞에서 언급했듯이 [그림 4]에서 보여주었던 $r > 0$ 일 때에 진화시간의 일정한 값으로서의 수렴은 개체수가 무한할 때의 현상이며, 개체수가 유한할 때에는 DNA길이가 임계DNA길이 이상 길어지면, DNA길이가 길어짐에 따라 진화시간이 지수적으로 증가한다는 것을 알 수 있다.

결합모델에서 개체수를 고정시키고 비대칭변수 r 를 변화시키면, 비대칭변수에 무관하

게 비슷한 DNA길이 L 에서 개체수가 유한할 때의 진화시간이 개체수가 무한할 때의 진화시간으로부터 벗어나서 지수적으로 증가하기 시작하는 것을 볼 수 있었다. 결합모델에서 개체수가 유한할 때에 진화시간이 개체수가 무한할 때의 진화시간으로부터 벗어나서 어느 정도 증가한 후에는, 비대칭변수에 무관한 증가율을 갖고 지수적으로 증가하는 것을 볼 수 있었다. 또한 결합모델에서 우리가 사용한 변수들의 값을 가정하고 비대칭변수를 10배 증가시키면 $t=10^6$ 세대 안에 바뀔 수 있는 뉴클레오티드의 최대 개수(L)가 대략 3개 증가하였으며, 비대칭변수가 100보다 커지면 비대칭변수가 더 증가하더라도 진화시간이 더 이상 감소하지 않고 바뀔 수 있는 뉴클레오티드의 최대 개수도 증가하지 않았다. 분리모델을 사용하여도 결합모델에서와 유사한 결과들을, 즉 DNA길이가 임계DNA길이 이상 길어지면 개체수가 유한할 때의 진화시간이 개체수가 무한할 때의 진화시간으로부터 벗어나서 DNA길이가 길어짐에 따라 지수적으로 증가하며 비대칭변수에 무관한 기울기를 갖는다는 것과 비대칭변수가 100 이상 커지면 비대칭변수가 더 증가하더라도 진화시간이 더 이상 감소하지 않고 바뀔 수 있는 뉴클레오티드의 최대 개수가 증가하지 않는다는 것을 얻을 수 있었다.

본 논문에서는 진화시간을 계산할 때에 전체 개체수의 1%가 대립유전자 S_0^* 를 갖는데 걸리는 시간을 구하였는데, 여기서 1%라는 수치는 큰 의미가 없다. 왜냐하면, 대립유전자 S_0^* 를 갖는 개체수가 $1/s$ 보다 크면, 개체수가 지속적으로 증가하여 반드시 안정적 상태에 도달하기 때문이다. 또한 DNA길이가 어느 정도 이상 긴 경우에는, 증가에 성공할 개체가 대립유전자 S_0^* 에 도달하는데 걸리는 시간이 도달한 개체가 증가하는데 걸리는 시간보다 훨씬 길기에, 선택기준을 0.1%, 1%, 10% 중 어떤 것으로 하든지 진화시간이 거의 같기 때문이다(Gill, 2010a: 195).

III. 요약 및 고찰

진화이론학자들이 진화를 설명하기 위하여 고안된 모델을 사용하여서 컴퓨터 시뮬레이션으로 진화시간을 계산한 최근의 연구 결과들을 살펴보았다. 여러 진화의 작용력 중에서 돌연변이와 자연선택을 고려한 돌연변이-자연선택 모델을 사용하였고, 단순지형에서 단성생식을 하는 반수체를 고려하였다. 돌연변이로는 염기치환을 고려하였고,

생식과정에서의 복제오류에 의한 돌연변이를 고려한 결합모델과 돌연변이원에 의해서 일생 중 어느 때라도 가능한 돌연변이를 고려한 분리모델 모두를 살펴보았다. 자연상태에서 발견되는 대립유전자인 야생형 S_0 에만 높은 적응도를 부여하고 나머지 대립유전자들에게는 낮은 적응도를 부여하는 단순지형에서는 결합모델과 분리모델 모두에서, 야생형 근처에 국소화된 유사종 형성과 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 야생형의 모든 뉴클레오티드가 치환된 대립유전자 S^*_0 에 야생형보다 큰 적응도를 부여하는 단순비대칭지형으로 바꾸었다. 야생형 근처에 국소화되었던 분포로부터 단순비대칭지형으로 바뀌고 난 후에 전체 개체수의 1%가 대립유전자 S^*_0 를 갖는데 걸리는 시간, 즉 진화시간을 DNA길이, 개체수, 비대칭변수 등을 변화시키며 계산하였다. 초기 대립유전자 분포의 야생형 근처에 국소화된 정도를 일정하게 유지하기 위하여 퍼짐변수의 값을 고정시키고 DNA길이를 변화시키면서 진화시간을 구하였다. 개체수가 무한하며 비대칭변수 $r=0$ 일 때에, 즉 야생형 S_0 의 적응도 W_0 와 대립유전자 S^*_0 의 적응도 W_L 이 같을 때에, DNA길이가 길어짐에 따라 진화시간이 지수적으로 증가하였다. 하지만, 개체수가 무한하며 비대칭변수 $r>0$ 일 때에, 즉 $W_L>W_0$ 일 때에, DNA길이가 어느 정도 이상 길어지면 DNA길이가 더 길어지더라도 진화시간은 증가하지 않고 거의 일정한 값을 가졌다. 개체수가 무한하여 비대칭변수가 너무 크지 않으면, 즉 $r\ll 1$ 이면, 결합모델과 분리모델의 진화시간은 거의 같았다. 개체수가 유한하면 비대칭변수 $r>0$ 일 때에도, DNA길이가 임계DNA길이 이상 길어지면 DNA길이가 길어짐에 따라 진화시간이 지수적으로 증가하였으며, 지수적인 증가율은 개체수와 비대칭변수에 무관하였고, 개체수가 무한하며 $r=0$ 일 때와 같은 증가율이었다. 또한 비대칭변수가 100 이상 커지면 비대칭변수가 더 증가하더라도 진화시간은 더 이상 감소하지 않았다. 위에서 언급한 개체수가 유한할 때의 결과들은 결합모델과 분리모델 모두에서 동일하게 얻을 수 있었다.

유전학에서 집단은 같은 종에 속하고 지리적으로 가까운 위치에 살아서 공통된 유전자집단인 유전자 풀(gene pool)을 가진 무리를 뜻한다. 우리는 그래프의 특성을 눈에 잘 띄게 하기 위하여, 집단의 개체수 $N=10^7, 10^{10}$ 이란 제법 큰 숫자를 선택하였다. 참

고로 대략적인 집단의 개체수를 살펴보면, 원핵생물은 $10^8 \sim 10^9$, 단세포진핵생물은 $10^7 \sim 10^8$, 무척추동물은 $10^5 \sim 10^6$, 척추동물은 $10^4 \sim 10^5$ 이다(Lynch and Conery, 2003). 돌연변이율을 $\mu=0.001$ 로 놓았는데, 생명체가 가지는 DNA 수선기능 때문에 실제 돌연변이율은 이보다 훨씬 작다. 참고로, 세균 유전자들의 전형적인 돌연변이율은 세포분열 당 $10^{-8} \sim 10^{-10}$ 이며, 진핵생물의 유전자 돌연변이율은 배우자(gamete)당 $10^{-5} \sim 10^{-6}$ 이다(Pierce, 2009: 477). 우리가 선택했던 돌연변이율보다 작은 값을 가정하면, 유전자변이가 생길 확률이 감소하여서 진화시간이 증가하게 된다.

우리는 돌연변이로서 DNA의 뉴클레오티드 하나가 바뀌는 염기치환을 고려하였으며, 뉴클레오티드 쌍이 첨가되는 삽입(insertion), 뉴클레오티드 쌍이 제거되는 결실(deletion) 등은 고려하지 않았다. 하지만, 자연계에 전혀 존재하지 않았던 새로운 유전정보를 가진 유전자가 만들어지기 위해서는, 삽입이나 결실의 과정이 포함될 수는 있지만, 결국 하나 하나의 뉴클레오티드가 우연히 조합되는 수밖에 없다고 본다. 비유로 말하면, 전혀 새로운 컴퓨터 프로그램이 우연히 만들어지기 위해서는 하나하나의 프로그램 코드가 우연히 조합되는 수밖에 없다. 마찬가지로, 주변의 DNA를 획득하는 형질전환(transformation)이나, 박테리오파지에 의해서 DNA가 옮겨지는 형질도입(transduction)이나, 플라스미드(plasmid) 등에 의해서 다른 세포의 DNA가 직접 전달되는 접합(conjugation)과 같은 수평적 유전자전달(horizontal gene transfer)에 의해서, 혹은 이동가능한 DNA 염기서열인 트랜스포존(transposon)이 DNA의 다른 위치에 삽입됨으로서, 자연계에 전혀 존재하지 않았던 새로운 유전정보를 가진 유전자가 만들어질 수 없다(Pierce, 2009). 또한 개체수가 유한할 때에 생기는 유전적부동은 컴퓨터 시뮬레이션으로 진화시간을 계산할 때에 이미 고려가 되었으며, 창시자 효과(founder effect)와 유전적 병목(genetic bottleneck) 등은 소수의 돌연변이체를 증폭시키는 효과는 있을지 몰라도 새로운 유전정보를 가진 유전자를 만들어 낼 수는 없다.

본 논문에서는 적응지형도로 단순지형을 고려하였는데, 야생형으로부터의 해밍거리가 증가함에 따라 적응도가 지수적으로 감소하는 배수지형(multiplicative landscape)에서도 단순지형에서 얻었던 결과들을 - DNA길이 임계DNA길이 이상 길고 개체수가 유한하면, DNA길이 길어짐에 따라 진화시간이 지수적으로 증가한다는 것과 비대칭변수가 100 이상 커지면 비대칭변수가 더 증가하더라도 진화시간이 더 이상 감소하지 않는다는 것을 - 동일하게 얻을 수 있었다.¹⁾ 따라서 본 논문에서 요약된 연구 결과의 핵심은 새

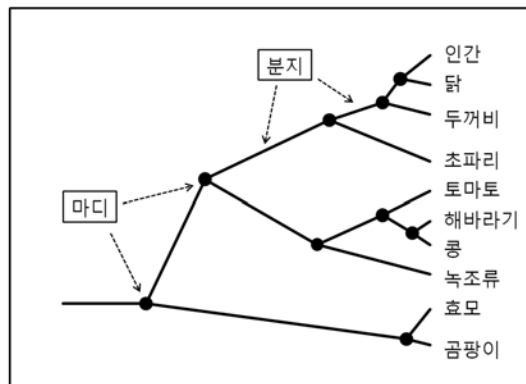
로운 유전정보를 가진 유전자를 만들기 위하여 바뀌어져야 할 뉴클레오티드의 개수가 증가함에 따라 진화시간이 지수적으로 증가하기에, 새로운 유전자가 아무리 큰 적응도를 가진다 하더라도 기존의 유전자로부터 어느 정도 이상의 뉴클레오티드가 바뀌어야만 한다면, 그 새로운 유전자로의 진화는 불가능하다는 것이다. 예를 들어서 [그림5]에서 DNA길이 L이 우리가 계산한 길이 이상으로 길어질 때에 진화시간이 같은 비율로 계속 지수적으로 증가한다고 가정한다면, 대장균의 경우에 최적 조건에서 한 세대가 약 20분 이기에 1억년($\sim 2.63 \times 10^{12}$ 세대) 동안 바뀔 수 있는 뉴클레오티드의 최대 개수(L)는 개체 수 $N=10^{10}$ 일 때에 대략 결합모델에서는 46개, 분리모델에서는 49개이다. 개체수 혹은 비대칭변수가 10배 증가할 때에 바뀔 수 있는 뉴클레오티드의 최대 개수는 대략 3개 증가 하며, 비대칭변수가 100보다 커지면 뉴클레오티드의 최대 개수가 더 이상 증가하지 않는다는 점을 감안하면, 자연계에서 존재가능한 개체수와 비대칭변수 등을 가정할 때에 1억년 동안 바뀔 수 있는 뉴클레오티드의 최대 개수는 대략 60개에 불과하고, 그 최대 개수보다 많은 뉴클레오티드가 바뀌어져야 하는 새로운 유전자로의 진화는 불가능하다.

여기서 첨가하고 싶은 점은 우리가 고려한 길이 L인 DNA는 새로운 기능을 가져서 높은 적응도를 가진 유전자가 되기 위하여 바뀌어져야 할 DNA 부분만을 지칭한다는 것이다. 예를 들어서 유전자가 1000 염기쌍으로 구성되어 있는데, 새로운 유전자가 되기 위하여 그 중에서 30 염기쌍만 바뀌면 된다고 할 때에, 우리가 계산한 DNA길이 L은 30이 되며, 본 논문에서는 30개의 염기쌍이 바뀌는데 걸리는 시간을 계산하였다. 하지만, 이러한 계산은 나머지 970개의 염기쌍이 그동안 바뀌지 않는다고 가정하지만, 이러한 가정은 실제로는 맞지 않다. 나머지 970개의 염기쌍도 비슷한 돌연변이율을 갖고 바뀔 가능성이 높기에, 1000개의 염기쌍 중에서 정확히 우리가 원하는 30개만 바뀌어서 새로운 유전자가 될 확률은 본 논문이 고려한 경우의 확률보다 훨씬 작으며, 따라서 실제 진화시간은 본 논문에서 제시한 값보다 훨씬 길어진다는 점을 염두에 두어야 한다.

골드(Gould)가 상당기간 정체(stasis) 상태에 있다가 어느 단계에서 짧은 시간에 급격하게 형태적 변화가 일어난다는 단속평형(punctuated equilibrium) 모델을 제안하였지만, 급속한 형태적 변화가 일어나려면 제법 많은 새로운 유전자들이 짧은 시간 동안에 생겨나야 하는데, 한 개의 새로운 유전자가 만들어지는데도 오랜 시간이 걸린다는 본 논문의 결과로 볼 때에 짧은 시간동안에 제법 많은 새로운 유전자가 만들어지는 것은 확률적으로 불가능하

1) 이 결과는 현재 학술지에 투고하기 위하여 논문 작성 중이다.

다(이양림, 1990). 오랫동안 점진적으로 변화하여서 새로운 종이 생겨났다는 계통진화적 점진주의(phyletic gradualism)가 본 논문의 결과로 볼 때에 단속평형 모델보다는 확률적으로 가능성이 높다고 볼 수 있다. 하지만 점진적으로 변화가 되었다면 수많은 중간단계의 생명체가 화석으로 발견되어야 하는데, 점진적인 변화를 나타내지 않는 화석 기록과는 부합하지 않는다는 지적을 받고 있다(Hyde, 2009). 또한 점진적인 변화과정으로는 설명하기 어려운 경우들이 자연계에 많이 있다. 예를 들어서, 폭격수땃정벌레는 적을 향하여 매우 뜨거운 퀴논이란 기체를 발사하기 위하여 과산화수소와 하이드로퀴논이란 화학물질, 평소에는 화학반응을 막는 억제물질, 폭격해야 할 때에 급격한 반응을 일으키게 하는 효소들, 100°C의 화학물질을 견디는 연소관, 기체를 만들고 발사하는 정확한 신경체계 등을 갖추고 있다. 땃정벌레로부터 폭격수땃정벌레로 진화하기 위하여 위에서 열거한 조건 중의 일부만 갖추면 아무런 소용이 없으며 오히려 땃정벌레 자신에게 위험하기에, 짧은 시간 동안에 위에 열거한 모든 조건을 갖추어야 한다. 그러나 이러한 조건을 갖추기 위하여 많은 새로운 유전자들을 짧은 시간 동안에 만드는 것은 확률적으로 불가능하다. 다른 예로서, 수정체, 망막 등을 갖춘 눈의 출현, 날개, 깃털, 공기저항이 적은 체형, 가벼운 뼈의 구조, 날개를 조작하는 두뇌 등을 갖춘 조류의 출현 등을 생각해 볼 수 있다.



[그림6] 생명체들의 유전자를 비교하여 만든 계통수의 예.

사실상 현재 지구상에 존재하는 다양한 종들 사이에는 굉장히 넓은 유전적 간격이 존재한다. 최근에 현재 생존하는 생명체들의 유전자들을 비교분석함으로써 [그림6]과 같은

계통수(phylogenetic tree)를 만드는데, [그림6]에서 마디(node)는 종분화(speciation)가 일어나기 전의 공통조상을 뜻하며, 마디에서 현재 생존하는 생명체까지의 분지(branch) 길이는 진화시간을 나타낸다. 그런데 계통수에 있는 수많은 마디에서 종분화가 일어났다면 마디에 해당하는 공통조상의 개체수는 상당히 많았어야 하는데 화석으로 발견되지 않으며, 또한 오랜 시간에 걸쳐서 공통조상으로부터 현재의 생명체까지 진화가 일어났다면 수많은 중간단계의 생명체가 존재했어야 하는데, 그런데도 불구하고 중간단계 생명체의 화석이 발견되지 않고 있다는 문제점이 지적되고 있다. 뿐만 아니라, 본 논문에서 요약 하였던 진화시간 결과에 따르면, 공통조상으로부터 현재의 생명체가 되기 위하여 갖추어야 할 모든 새로운 유전자들을, 진화론에서 고려하는 작용력들에 의해서 진화론자들이 제시하는 진화시간 이내에 갖추는 것은 불가능하다고 본다.

요약을 하면 단속평형 모델은 수많은 유전자가 짧은 시간에 생겨나야 하기에 확률적으로 불가능하며, 반면에 계통진화적 점진주의는 유전자들이 단계적으로 생겨나는 것을 허용하기에 단속평형 모델에 비해서 상대적으로 가능성이 높다. 그렇지만, 진화론에서 제시하고 있는 작용력들에 의한 DNA의 염기쌍 변화가 매우 느리다는 본 논문의 결과로부터 유추해 볼 때에, 원시생명체로부터 유전자들이 점진적으로 변화하여서 현재의 다양한 모든 종이 생겨났다는 논리는 확률적으로 불가능하기에 설득력이 없다고 본다. 결론적으로 말하면, 진화론자들이 돌연변이, 자연선택, 격리 등을 사용하여 원시생명체에서 인간이 생겨났다는 생물학적 진화론을 설명하고자 노력하지만, 받아들이지 않으면 안 될 정도로 설득력있는 과학이론은 아니다. 앞으로 과학이 더 발전하면 축적된 대립유전자 빈도분포에 대한 자료로부터 적응지형도를 유추하고 진화시간을 좀더 정확히 계산할 수 있을 것으로 기대하며, 이와 같은 결과들을 통하여 더 명확히 진화가 불가능하다는 것이 밝혀지리라 기대한다. 그러기 위해서는 무신론자에 의해서 과학지식이 독점되어서 편견에 의해 왜곡되게 허용해서는 안 되며, 믿음과 실력을 겸비한 크리스천 학자들에 의한 과학지식에 대한 바른 해석이 꼭 필요하다고 본다.

“이 논문은 다른 학술지 또는 간행물에 게재되었거나 게재 신청되지 않았음을 확인함.”

참고문헌

- 이양림 (1990). 『진화론의 실상과 창조』. 서울: 생명의 말씀사.
- Baake, E. and W. Gabriel (2000). "Biological evolution through mutation, selection, and drift: An introductory review." in: Stauffer, D. (Ed.) (2000). *Annual Reviews of Computational Physics VII*. Singapore: World Scientific. 203-264.
- Baake, E. and T. Wiehe (1997). "Bifurcations in haploid and diploid sequence space models." *Journal of Mathematical Biology* 35. 321-343.
- Bürger, R. (1998). "Mathematical properties of mutation-selection models." *Genetica* 102/103. 279-298.
- Dawkins, R. (2007). *The God Delusion*. 이한음역 (2007). 『만들어진 신』. 서울: 김영사.
- Drossel, B. (2001). "Biological evolution and statistical physics." *Advances in Physics* 50(2). 209-295.
- Eigen, M., J. McCaskill and P. Schuster (1988). "Molecular Quasi-Species." *J. Phys. Chem.* 92. 6881-6891.
- Franz, S. and L. Peliti (1997). "Error threshold in simple landscapes." *J. Phys. A: Math. Gen.* 30(13). 4481-4487.
- Gill, W. (2010a). "Dependence of the Crossing Time on the Sequence Length in the Wright-Fisher Multiple Allele Model." *Journal of the Korean Physical Society* 57(1). 192-199.
- _____ (2010b). "Dependence of the Crossing Time on the Sequence Length in the Continuous-time Mutation-selection Model." *Journal of the Korean Physical Society* 57(2). 287-295.
- Higgs, P. G. (1994). "Error thresholds and stationary mutant distributions in multi-locus diploid genetics models." *Genet. Res. Camb.* 63. 63-78.
- Hofbauer, J. (1985). "The selection mutation equation." *Journal of Mathematical Biology* 23. 41-53.
- Hyde, D. (2009). *Introduction to Genetic Principles*. 김세재 외 역 (2010). 『대학생을 위한 유전학』. 서울: 라이프사이언스.
- Lynch, M. and J. S. Conery (2003). "The origins of Genome Complexity." *Science* 302. 1401-1404.
- Pierce, B. A. (2009). *Genetics-A Conceptual Approach*. (3rd ed.). 전상학 외 역 (2009). 『유전학의 이해 개념과 원리 제3판』. 서울: 라이프사이언스.
- Wagner, H., E. Baake and T. Gerisch (1998). "Ising Quantum Chain and Sequence Evolution." *Journal of Statistical Physics* 92. 1017-1052.

ABSTRACT

Summary and Discussion about the Recent Simulation Result in the Evolution Model

Won-Pyong Gill(Pusan National University)

We have summarized and discussed the recent computer simulation results in the mutation-selection model, which was designed to explain the evolution theory. The evolution time for a haploid, asexual population in the coupled and the decoupled mutation-selection model was calculated for various DNA lengths, population sizes and asymmetric parameters, by switching on an asymmetric sharply-peaked landscape. The evolution time for a finite population with a positive asymmetric parameter was found to be an exponentially increasing function of the DNA length whose rate was relatively constant, even though the population size and asymmetric parameter were varied with a fixed extension parameter. This suggests that the evolution to a new allele, which is required by many mutations whose number exceeds a certain limit, cannot be allowed, even though the new allele has an extremely large selective advantage. In conclusion, the evolution theory has no compelling evidence in spite of many efforts to explain by using various mechanisms, such as mutation, selection, and so on.

Key Words: evolution time, mutation-selection model, simulation, sharply-peaked landscape, biological evolution theory