

전위인자가 표식인자로 사용될때 발생하는 동시형질도입의 비대칭성

김창훈 · 박찬규*

한국과학기술원 생명과학과

선택적 표식인자가 전위인자일때, 표식인자의 선택 순서에 따라 동시형질도입빈도가 달라진다. 본 연구에서는 이 현상을 설명하기 위하여 아래와 같은 수학적 접근을 시도하였다. 먼저, transducing particle들의 형성과정을 다섯 종류로 유별하였다. 그다음, 하나의 표지가 고정될 가능성을 의미하는 확률함수들을 확률밀도 개념에 근거하여 수학적으로 정식화 하였다. 그리고 실제 값들과 유용한 가정들을 이 식들에 결합하여 그로부터 얻어진 이론적인 빈도를 실제 값들과 비교 하였다. 이러한 분석에 의하여 그 빈도의 상이성이 일반 재조합에 필요한 상동성을 전위인자가 제공하지 못하고, 그 주변에 재조합 빈도를 거리에 의존하는 형태로 감소시키기 때문이라는 결론을 얻었다.

KEY WORDS □ Cotransduction frequency, Transposon, Asymmetry, Homologous recombination

미생물 유전학에서 파아지를 이용한 형질도입(Transduction)은 유전자 지도 그리기 및 균주 만들기에서 있어서 매우 중요한 방법의 하나이다(3). 선택적 표식인자(Selectable marker)는 가장 기본적인 개념으로서 이 과정의 조작을 용이하게 한다. 표식인자는 그 성질에 따라 선택 가능한 표현형을 갖는 것과 갖지 못하는 것으로 구분해 볼 수 있다. 후자의 경우는 그 표식인자를 선별하는 데 있어서 많은 수고가 필요하였다. 이러한 난관을 극복하기 위해서 인접한 선택 가능 표식인자를 종종 이용하는데, 전위인자(Transposon)는 이 목적에 아주 적절하였다(1,2).

그런데, Tn10을 선택적 표식인자로 이용한 형질도입시에 표식인자의 선택 순서에따라 두 표식인자 사이의 동시형질도입(Cotransduction) 빈도가 달라짐이 발견되었다(2). 그 이유에 대해서는 Tn10 자체의 특이성 때문에 Transducing particle의 DNA 단편이 형성 될 때 Tn10의 한쪽 끝을 포함할 가능성이 높기 때문이라는 막연한 추측만 제시된 바 있다.

본 연구에서는 이 문제를 수학적으로 접근하여 Tn10이 재조합에 필요한 상동성(Homology)을 제공하지 못 할 뿐만아니라 그 주변에 재조합(recombination)이 일어나는 빈도를 감소시키는 데에 그 원인이 있을 가능성을 검토 하였다.

수학적 공식화

*E. coli*의 P1 일반 형질도입에서 대장균 염색체의 임의의 단편이 Transducing particle을 형성할 가능성은 동일하다고 밝혀져 있으며, 일반 재조합(General recombination)이 Particle 에 의해 세포 내에 들어온

DNA 단편(fragment)을 염색체에 끼워넣는데 주 역할을 하는 것으로 생각되고 있다.

이 두가지 사실을 전제하고, 표식인자의 선택순서에 따른 기준점의 변화, 전위인자를 지나는 점에서 합수의 불연속성, 그리고 각 표식인자의 포함여부 등에따라 Transducing particle들을 분류하고 이들 각각이 고정될 가능성을 고려하여 동시형질도입 빈도를 이론적으로 구해 볼 수 있다. 그림 1에서 처럼 각 단편에 대해서 선(先)선택적 표식인자(Selective marker)를 기준으로 하여 변수 및 상수를 정하고 이들간의 상관관계를 구할 수 있다.

그림 1에서 α 는 전위인자의 크기를 나타내는데 그 주변에 미칠 영향을 고려하여, α 의 한쪽 끝점으로부터 임의의 점까지의 거리를 σ 로 두고 σ 에서 재조합이 일어날 확률밀도함수 $r(\sigma)$ 를 정의하면, α 로부터 x 만큼 떨어진 점과 α 사이에서 재조합이 일어날 확률은 식(1)이 된다.

$$f(x) = \int_0^x r(\sigma) d\sigma \tag{1}$$

그림 1에서 어느 부분이 고정되기 위해서는 그 부분의 양쪽에서 적어도 각각 홀수 번의 재조합이 일어나야 한다. 사실상, 각 한 번씩 재조합이 일어나는 경우만을 고려하여도 무리는 없을 것으로 생각된다. 먼저, 그림 1의 a)의경우에 대해서 식 (1)에 따라 각 구간의 재조합율을 구하면,

$$\int_0^x r(\sigma) d\sigma \tag{2}$$

$$\int_{\beta}^{L-x-\alpha} r(\sigma) d\sigma \tag{3}$$

*본 논문에 대한 문의는 이 저자에게로.

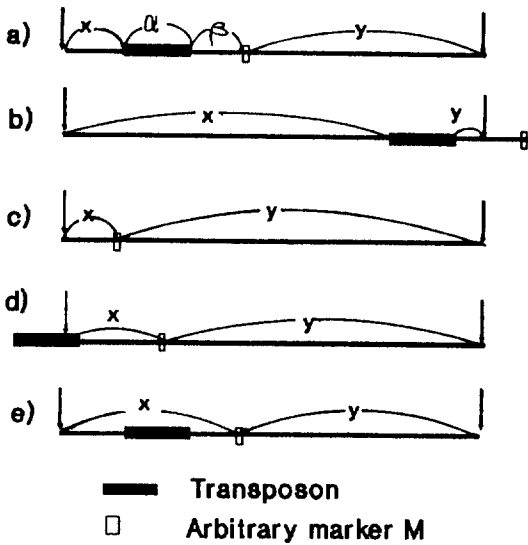


Fig. 1. Configurations chromosomal DNA fragment in size of 90 kb contained in transducing particles. The α is the size of transposon, β the distance between the right end of the transposon and the other marker M ; x , the distance to the left end of P1 genome from the anchoring point; y , the remaining part of the fragment. In case a), the region covering α and β does not allow recombination to occur, and the two markers are to be fixed together. Case b) shows that the transposon is used as a selective marker and that the unselective marker does not have to be fixed simultaneously. Case c), d), e) show that the marker M is used as a selective marker and that it is not necessary for transposon to be fixed together. Because of discontinuity in x domain, three intervals are considered independently. The anchoring point in the first two cases (case a, b) are in the left end of transposon, in other three cases (case c, d, e), the marker M . See text for details.

식 (2)와 식 (3)이 되고, 이 두 식을 곱하여 식 (4)가 구해진다($L = P1$ 의 genome 크기, $\beta = \alpha$ 의 왼쪽 끝으로부터 표식인자 M 까지의 거리). 여기에서, $F_1(x)$ 는 그림 1의 a)에서 α 와 β 부분을 모두 포함하여 고정될 확률을 나타낸다. b)의 경우도 위와 같이 하여 식 (5)가 된다($d = \alpha + \beta$). 식 (4)와 식 (5)에서 적분한계만 다를 뿐인데, 이 차이는 β 부분에서 재조합을 허용하느냐 하지 않느냐에 따른 것이다.

$$F_1(x) = \int_0^x r(\sigma) d\sigma \int_{\beta}^{L-x-\alpha} r(\sigma) d\sigma \quad 0 < x \leq L-d \quad (4)$$

$$F_2(x) = \int_0^x r(\sigma) d\sigma \int_0^{L-x-\alpha} r(\sigma) d\sigma \quad 0 < x \leq L-\alpha \quad (5)$$

식 (4), (5)는 α 의 왼쪽 끝으로부터 x 를 잡아서 계산하였는데, 식 (6), (7), (8)에서는 편의상 M (동시 도입되는 다른 표식인자)을 기준으로 x 를 취했다. 그림 1의 c 는 α 를 포함하지 않으므로 $r(\sigma) = k$ 가 됨 M 이 고정될 확률은 간단히 식 (6)이 된다.

$$F_3(x) = k^2 x(L-x) \quad 0 < x \leq \beta \quad (6)$$

d)는 형질도입 DNA의 한쪽 끝이 전위인자 내에 있는 다소 특수한 상황을 보여주는데, 상동성이 없는 부분의 길이가 변하고 있기 때문에 상당히 복잡한 간섭 효과가 예상되지만 그 길이자체가 $r(\sigma)$ 에 거의 영향을 미치지 않는다고 가정하면 이때 M 이 고정될 확률은 식 (7)과 같이 된다.

$$F_4(x) = \int_0^{\beta} r(\sigma) d\sigma \int_{\beta}^{L-x+\beta} r(\sigma) d\sigma \quad \beta < x \leq d \quad (7)$$

e)에서는 x 부분에서 전위인자의 좌측 및 우측 영역, 그리고 y 부분 이렇게 세 구간에서 재조합을 각각 구하여 x 부분의 율과 y 부분의 율을 곱하면 e)에서 M 이 고정될 확률이 되고, 식 (8)과 같이 된다.

$$F_5(x) = \left\{ \int_0^{\beta} r(\sigma) d\sigma \int_0^{x-d} r(\sigma) d\sigma \right\} \int_{\beta}^{L-x-\beta} r(\sigma) d\sigma \quad d < x \leq L \quad (8)$$

이상과 같이 각각의 경우에 대한 확률 $F_1 \sim F_5$ 를 구했으므로, 각 식 옆에 표시된 구간내에서 각 함수를 적분하여 아래와 같이 동시형질도입 빈도의 정의에 따라 조합하면 식(9), (10)과 같이 된다. 두 표식인자가 함께 고정될 가능성 (a)의 경우)을 α 가 고정될 가능성 (b)의 경우, 즉 적어도 α 가 포함될 가능성)으로 나누면 전위인자가 표식인자로 선택될 경우의 동시형질도입 빈도 C_1 이 되고 M 이 고정될 가능성(그림 1의 c), d), e)의 경우, 이 때도 이 표식인자가 적어도 포함될 가능성)으로 나누면 M 이 표식인자로 선택될 때의 동시형질도입 빈도 C_2 가 된다. C_1, C_2 의 정의 식에 $F_1 \sim F_5$ 의 적분항이 포함 된 이유는 함수 $F_1 \sim F_5$ 가 각 구간 내의 대표적인 예에 불과하므로, 그 구간에 속하는 것들 모두를 고려해야 하기 때문이다.

$$C_1(d) = \frac{\int_0^{L-d} F_1(x) dx}{\int_0^{L-\alpha} F_2(x) dx} \quad (9)$$

$$C_2(d) = \frac{\int_d^L F_1(x-d) dx}{\int_0^{\beta} F_3(x) dx + \int_{\beta}^d F_4(x) dx + \int_d^L F_5(x) dx} \quad (10)$$

비교분석

위에서와 같이 정식화된 일반적인 관계식(식 9, 10)이 동시형질도입 빈도의 비대칭성에 대한 설명을 가

Table 1. Comparisons between theoretical and actual cotransduction frequencies

d (Kb)	W (%) Reference ^a	C ₁ (%)		C ₂ (%)	
		Theoretical	Actual ^a	Theoretical	Actual ^a
0	100 ^b	100	*	100	*
0.1	99.7 ^b	99.6	*	99.6	*
*	*	*	95	*	83
4	87	85.6	94	82.6	42
*	*	*	75	*	34
6.4	80	78	67	73.9	31
*	*	*	60	*	15
40	17.1	12.8	*	10	*

^afrom Kleckner *et al* (2), ^bcalculated from d. *: Not applicable

시화 하지 못 하므로, 몇 가지의 가정을 도입하여 이론적으로 값들을 구하고 실제의 데이터와 비교 분석하므로써 그 비대칭성의 원인을 해석할 수 있다. 다음과같이 세 가지의 경우로 나누어 분석하는 것이 타당할 것으로 보인다.

1) $\alpha=0, r(\sigma)=k(\text{상수})$ 인 경우

이 값들을 식 (4)~(8)에 각각 대입하고 정리하면 $F_1 \sim F_5$ 가 구해진다. $\alpha=0$ 이고 $\beta=d$ 가 되므로 다음과 같이 된다.

$$F_1(X) = k^2 X(L - X - d) \tag{11}$$

$$F_2(X) = F_3(X) = F_5(X) = k^2 X(L - X) \tag{12}$$

$$F_4(X) = k^2 d(L - X) \tag{13}$$

정의식 9에 F_1, F_2 를 대입하여 각 구간에서 적분하면 $C_1 = (1 - d/L)^3$ 이 된다. 정의식 10에 F_1, F_3, F_4, F_5 를 대입하여 각 구간에서 적분하면 $C_2 = (1 - d/L)^3$ 이 되어, $C_1 = C_2$ 가 되면서 Wu의 식(5)이됨을 알수있다. 이때에는 동시형질도입 빈도의 비대칭성이 나타나지 않고 있는데, 애초의 가정들을 생각해 보면 당연한 귀결이라 하겠다.

2) $\alpha = \text{일정}, r(\sigma) = k(\text{상수})$ 인 경우.

위에서와 마찬가지로 식(4)~식(8)에 각 값을 대입하고 정리하면 $F_1 \sim F_5$ 를 구하게 되고, 이를 식 (9), (10)에 다시 대입하면 C_1, C_2 가 구해진다. $\alpha=0$ 인 경우에는 이미 $F_1 \sim F_5$ 를 제시 했으므로 여기서는 적분한 형태를 제시한다. 실제의 함수값은 다를지라도 원리는 동일하므로 중복을 피하기 위해서이다. 물론, 식 (14)~식(18)에 $\alpha=0, \beta=d$ 를 넣으면 1)의 경우에서의 적분식들과 일치한다.

$$\int_0^{L-d} F_1(x) dx = k^2(L-d)^3/6 \tag{14}$$

$$\int_0^{L-\alpha} F_2(x) dx = k^2(L-\alpha)^3/6 \tag{15}$$

$$\int_0^\beta F_3(x) dx = \int_0^\beta k^2 x(L-x) = k^2 \beta^2(L/2 - \beta/3) \tag{16}$$

$$\int_\beta^d F_4(x) dx = \int_\beta^d k^2 \beta(L-x) = k^2 \beta(d-\beta)(L-d/2 - \beta/2) \tag{17}$$

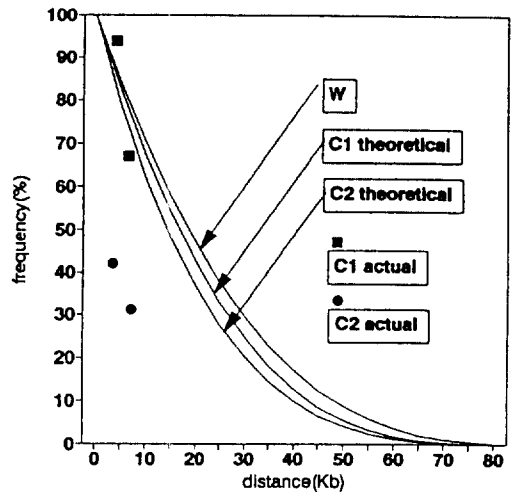


Fig. 2. Comparisons between theoretical and actual cotransduction frequencies, where $\alpha=9.3$ kb, $L=90$ kb, $r(\sigma)=k(\text{constant})$. See table 1 for the values. Distance between the two markers is β without Tn10, and insertion of Tn10 increase the distance by α . Thus, the in W and $\alpha + \beta$ in C_1/C_2 are regarded as a distance criterion for comparisons. See text for details.

$$\int_a^L F_5(x) dx = k^2(L-d)^2(L+2d-3\alpha)/6 \tag{18}$$

그러나, 여기서는 1)의 경우처럼 간단하지는 않다. 표 1에 $\alpha=9.3$ Kb (Tn10의 실제 크기), $L=90$ Kb(P1의 genome 크기)일때 이론적 계산치를 구하여 그 값들을 실제 데이터(2)와 함께 나타내었다. 이 결과에서 우리는 다음과같은 사실을 알 수있다(표1, 그림2 참조).

- a) 전 구간에서 $W > C_1 > C_2$ 이다(단, W는 Wu의 식에의한 동시형질도입 빈도).
- b) 실제 데이터에서는 이론치의 감소 경향을 유지하나 일반적으로 더 큰 차이를 보인다.
- c) 실제의 데이터는 α 의 크기를 더 크게한 ($\alpha=50$)

경우의 이론치와 유사하다.

3) $\alpha = \text{일정}$, $r(\sigma) = \sigma$ 의 함수인 경우

a), b), c)에서 알 수있듯이 2)의 가정만으로는 설명이 충분치 않으므로, $r(\sigma)$ 가 상수가 아닌 경우를 고려함으로써 보완하고자 한다. 그러나, $r(\sigma)$ 가 어떠한 함수인지 알 수없기 때문에, 이 함수가 어떤 성질을 가질 것인가를 알아보고, 그 성질을 바탕으로 하여 실제 데이터와의 일치 가능성을 논해 볼 것이다.

먼저 c)에서, 이 표현을 ' α 의 주변에 재조합율이 감소하였다.'라고 가정한다면 비슷한 결론이 이끌어 짐을 알 수 있다. 재조합율의 감소 정도는 σ 가 감소함에 따라 증가하여 $\sigma = 0$ 이면 재조합율은 0이 되고, $\sigma = \infty$ (매우크다)이면 재조합율은 전위인자가 없는 상태와 동일하게 된다. $\sigma_1 < \sigma_2$ 이면 $r(\sigma_1) < r(\sigma_2)$ 가 된다.

이제, C_1 과 C_2 를 비교하면 F_3 항은 간섭효과를 받지 않으며 분자는 서로 같으므로, 간섭의 정도가 클수록 C_1/C_2 가 더 커지게 된다. 이것은 부분적인 설명이지만 바람직한 결과이다. σ 가 매우 작은 값을 가지면 두 점 사이에서 재조합이 일어날 가능성은 간섭이 없는 경우보다 줄어든다. 즉, 간섭 효과가 크면 두 표식인자가 함께 도입될 가능성이 커진다. 이 결론은 a)와는 다르다. 왜냐하면, $C_1, C_2 > W$ 혹은 $C_1 > W > C_2$ 가 될 수도 있음을 나타내기 때문이다.

이렇게 수정된 우리의 가정이 두 번째의 가정보다 실제의 데이터를 더 잘 설명하므로 가정이 타당함을 알 수 있다.

고 찰

전위인자의 특징으로서 동시형질도입 빈도의 비대칭성을 설명하는 것은 타당성을 지닌다. 그러나, 전위인자의 특성에만 집착하다 보면 현상의 배후에 놓여 있는 더 근본적인 원인을 파악하기 어렵다.

우리는 동시형질도입 빈도의 비대칭성이 전위인자의 특이한 성질에 기인하는 것이 아니라, 전위인자가 재조합에 필요한 상동성을 제공하지 못하기 때문일 것이라는 가정에서 출발하였으나, 수학적 분석의 결과는 단순히 그 이유만이 아님을 보여주었다. 즉, 동시형질도입 빈도의 비대칭성은 상동성이 없는 부분과 그것의 간섭효과 때문일 것이라는 추측이다. 이것은 상당히 일반적인 설명으로서, 전위인자가 아닌 다른 DNA 단편의 삽입(insertion)만으로도 동일한 효과를

기대할 수 있다는 말이 된다. 그리고 이러한 경우들에 있어서 간섭효과가 구배(gradient)를 갖는다는 가정이 오히려 타당한지도 모른다.

일반 재조합에서 상동성이 없는 부분이 Looping-out되는 것으로 생각 되는데 이로인한 입체장애효과도 있을 것으로 보인다. 그 밖에 현재까지 알려지지 않은 재조합기작과 관련된 더 근본적 기작도 관련될 수 있다.

결국, 어떠한 기작에 의하든, 서로 다른 두 선택적 표식인자의 선택 순서에따른 동시형질도입 빈도의 비대칭성은 각 표식인자의 형질도입 빈도가 달라진 데에 원인이 있는 것이다. 실험과 관련하여 정의된 동시형질도입 빈도는 두 표식인자의 동시도입 빈도를 선택적 표식인자의 형질도입 빈도로 정규화 한 것이어서, 엄밀한 의미에서 동시도입 빈도에는 비대칭성이 나타나지 않음에도 불구하고, 정규화 과정에서의 기준 차이로 말미암아 비대칭성이 관찰 되는 것이다. 따라서, 동시형질도입 빈도의 개념을 실험적 편의성보다 개념적 정확성에 초점을 두고 정의 한다면 간섭효과와 결과는 단순히 두 표식인자 사이의 거리가 가까워 지는것으로 나타날 것이다.

사 사

본 연구는 부분적으로 한국과학재단의 목적기초 연구비의 지원에 의해 이루어졌음을 밝힌다.

참 고 문 헌

1. Davis, R.W., D. Botstein, J.R. Roth, 1980. Advanced bacterial genetics: a manual for genetic engineering. Cold Spring Harbor.
2. Kleckner, N., D. Botstein, J. R. Roth, 1977. Genetic engineering *in vivo* using Translocatable drug-resistance elements. *J. Mol. Biol.* **116**, 125-159.
3. Miller, J.H., 1972. Experiments in molecular genetics. Cold Spring Harbor.
4. Scaife, J., D. Leach., A. Galizz., 1985. Genetics of bacteria. Academic press, INC: 197-215.
5. Wu, T.T., 1966. A model for three-point analysis of random general transduction. *Genetics*, **54**, 405-410.

(Received January 29, 1992)

(Accepted March 11, 1992)

ABSTRACT: Asymmetry of Cotransduction Frequency Occurred When the Transposon Is Used as a Marker

Kim, Changhoon and Chankyu Park (Department of Life Science, KAIST.)

When the transposon is used as a selectable marker, the cotransduction frequency depends on the selection order of the markers. In this study, we adopted a mathematical approach to explain this phenomenon. At first, the formation of transducing particles were considered in five different configurations. Then the probability functions indicating the possibilities for a marker to be fixed were mathematically formulated on the basis of probability density concept. After actual values and useful assumptions were integrated into the formula, resulting frequencies from theoretical calculations were compared with actual data. Such analysis let us conclude that the asymmetry in the frequency arose from the lack of homology required for homologous recombination due to the transposon insertion and from the suppression of recombination around the region where the transposon is inserted.