

Hanwoo individual identification with DNA marker information

Jea-Young Lee¹⁾ · Yu-Mi Choi²⁾

Abstract

This study was conducted to establish an individual identification system in Hanwoo cattle. Samples of 33 Hanwoo individuals from Korean elite sire families were used. Thirteen major microsatellite markers were selected from alleles amplified, their frequencies, H(Heterozygosity) and PIC(Polymorphism Information Content) with Hardy-Weinberg equilibrium. Next, in order to evaluate the power of the markers selected on the individual animal identification, MP(Match probability) and R(Relatedness coefficient) with the percentage of animal incorrectly identified were computed. Finally nine microsatellite markers were selected and discussed.

Keywords : Heterozygosity, Match Probability, Microsatellite Marker, Polymorphism Information Content, Relatedness Coefficient

1. 서론

최근 미국, 일본 등지에서 광우병이 발생하고 국내에서도 구제역이 발생하는 등 안전한 축산물 공급을 저해하는 요인들이 증가하고 있다. 이에 대한 한 가지 대처 방안으로 한우생산이력제 구축을 들 수 있겠다. 한우생산이력제를 시행하면 식품안전 문제가 발생하거나 젓소고기가 한우로 둔갑 판매되는 행위에 신속히 대처 할 수 있을 뿐만 아니라 생산에서 소비에 이르는 유통과정 상의 신뢰성을 높일 수 있을 것이다. (Lim 등, 2005) 이에 기존의 개체의 동일성 검사에 사용된 방법으로 혈액형 및 생화학적인 지표의 다형성 분석방법을 사용하였으나 소에 있어서 혈액형검사는 정확한 부계혈통을 결정하지 못하는 단점이 있다(Glowatzki-Mullis 등, 1995). 이러한 단점을 해결하기위해 DNA 다형성에 기초한 다양한 유전분석 방법들이 개발되었으며, 이를 이용하여 품종간 및 개체간의 관별이 이루어지고 있다(Peelman 등, 1998; Arena 등,

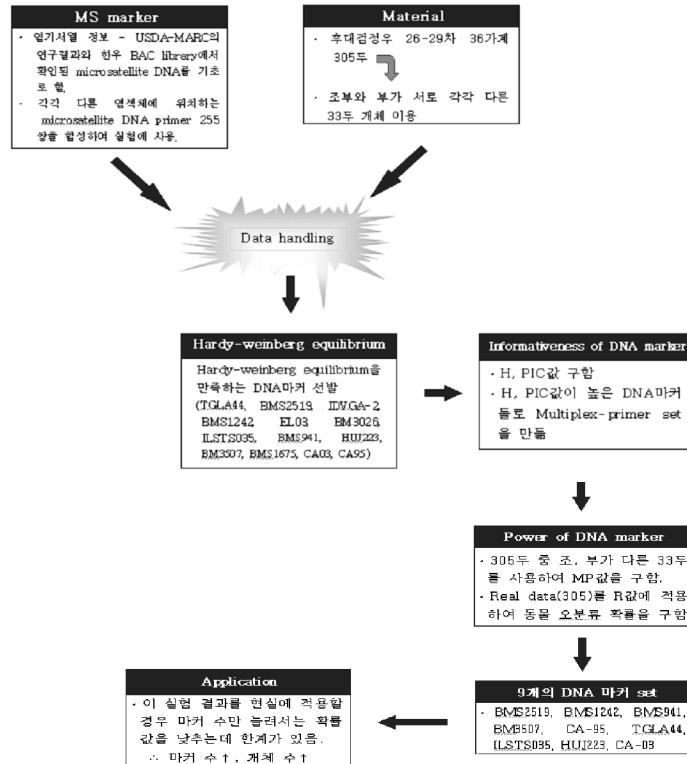
1) Professor, Department of Statistics, Yeungnam University, Kyungsan, Korea
Correspondence : jlee@yu.ac.kr

2) Graduate, Department of Statistics, Yeungnam University, Kyungsan, Korea

2002; Yeo 등, 2006). 위와 같은 현상을 현실에 적용하기 위해 Gyeongbuk Hanwoo cluster(GBHC, 경북한우클러스터)에서는 경상북도 산·학·연·관이 협력체계를 구축하고 DNA 유전자 분석을 통해 고품질의 한우를 생산하기 위해 노력하고 있으며, 실험한 실제 한우의 microsatellite marker 데이터를 본 논문에 활용하고자 한다. 본 논문은 사전 논문(Yeo 등, 2006)에서 5개의 마커를 사용하여 개체식별에 유용한 마커 셋을 찾았는데 이에 마커를 더 늘렸을 때 개체 식별 정확도가 더 높아질 것인가를 고려하게 되었다. 그러나 마커 수만 늘렸을 경우 개체 식별 정확도는 높아지지만 실험 비용, 시간이 많이 들므로 실용성이 떨어질 것으로 보여 진다. 따라서 마커 수를 5개에서 9개로 늘렸을 때 마커 정보력의 척도인 H(Heterozygosity)와 PIC(Polymorphism Information Content) 통계량을 이용하여 개체 식별에 가장 유용한 마커를 찾아내고 이 선택된 마커를 하나의 셋으로 만들어 이 셋의 검정력을 MP(Match probability)와 R(Relatedness coefficient)계수를 통해 알아보고, 선택된 셋을 적용하여 최종적으로 개체식별에 유용한 DNA 마커 셋을 찾아 그 결과를 현실에 적용해보고자 한다.

2. 실험방법 및 재료

본 연구에 공시된 한우는 농협중앙회 가축개량 사업소로부터 한우집단의 혈연정보가 정확히 기재된 국가 후대검정우 26-29차 36가계 305두와 그 중 조부와 부가 서로 각각 다른 33두 개체를 이용하였고, 통계적 분석과정은 <그림 1>의 과정으로 이루어졌다.



<그림 1> 개체식별에 유용한 DNA 마커를 찾는 과정

위 과정은 국가 후대 검정우 26-29차 36가계 33두의 염색체에서 155개의 DNA 마커를 이용하여 Hardy-Weinberg equilibrium을 만족하고 H값이 높은 13개의 마커를 선발하였고 이 마커를 이용하여 DNA마커의 검정력을 평가해 개체식별에 가장 유용한 조합을 찾았으며, 그 결과를 현실에 적용하는 과정을 나타내었다.

3. DNA marker 정보력의 측도

DNA 마커의 정보력이란 개체식별의 측면에서 보았을 때 다양성의 측도라고 할 수 있는데 이는 다양한 종류의 alleles를 가지고 있는 DNA 마커가 개체식별에 더 유용한 것을 뜻하며, 정보력의 측도로 이형접합체(H)와 다양성 정보량(PIC)을 사용한다. H, PIC값은 DNA 마커가 얼마나 많은 allele들이 균일하게 퍼져있는가를 나타내는 척도라 할 수 있고, 1에 가까울수록 개체식별에 유용한 DNA 마커라고 할 수 있다 (Arana et al., 2002; Ott, 1999).

3.1 이형접합체 측도 (Arana et al., 2002; Ott, 1999)

$$H = (1 - \sum_{i=1}^n p_i^2) \quad (1)$$

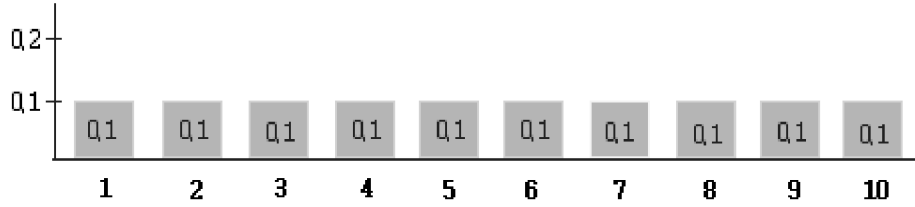
단, n = locus의 allele의 수
 p_i = locus i 번째 allele의 추정된 도수 비율

그러나 이형접합체(Heterozygosity:H)는 대립 유전형질의 독립성 보장을 만족하는지를 검정하기 위한 Hardy-Weinberg equilibrium 조건을 검정 하여야한다. 따라서 먼저 Hardy-Weinberg equilibrium검정(외적인 요인이 작용하지 않는다면, 유전자와 빈도 모두가 변하지 않고 평형을 이루게 되는 것)을 먼저 실시해 보았다. 유전자좌 내에서의 독립성 검정은 χ^2 통계량을 이용한 적합도 검정으로 각 유전자형 범주에 대한 실제 관측빈도와 기대빈도를 이용하여 구한다. 이를 국가 후대 검정우 26-29차 36가계 조부와 부가 서로 다른 33두 개체를 이용하여 구해본 결과 TGLA44, BMS2519, IDVGA-2, BMS1242, EL03, BM3026, ILSTS035, BMS941, HUJ223, BM3507, BMS1675, HW-YU2-CA03, HW-YU2-CA95이 하디-와인버그 평형검정을 만족하는 마커로 분석에 이용하였다.

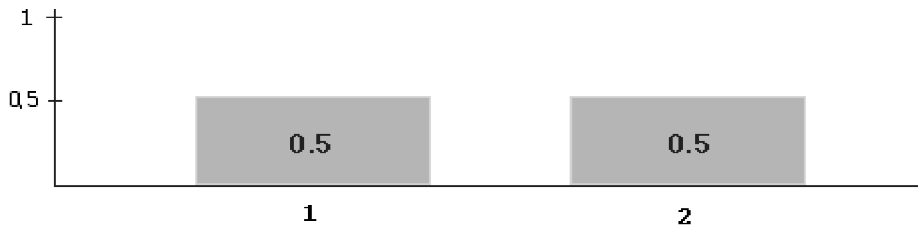
3.2 다양성 정보량 측도 (Weiss, 1993)

$$PIC = (1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2) \quad (2)$$

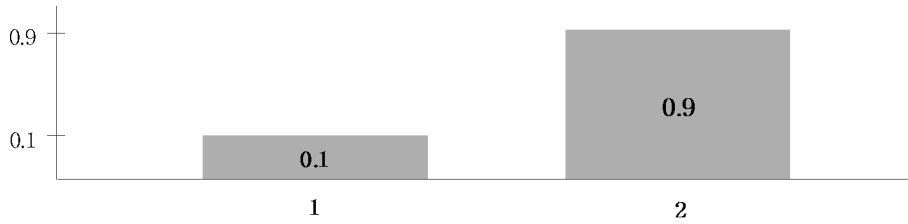
where n = locus의 allele의 수
 p_i = locus i 번째 allele의 추정된 도수 비율
 p_j = locus j 번째 allele의 추정된 도수 비율



<그림 2> $H=0.9$, $PIC=0.89$



<그림 3> $H=0.5$, $PIC=0.38$



<그림 4> $H=0.18$, $PIC=0.163$

위의 <그림 2>와 <그림 3>의 경우 <그림 2>는 10개의 allele과 0.1의 균일한 allele 도수를 가진 마커이고, <그림 3>은 2개의 allele과 0.5의 균일한 allele 도수를 가진 마커이다. 이 두 마커는 allele 도수는 균일하지만 <그림 3>은 <그림 2>에 비해 다양한 종류의 allele을 가지고 있지 않기 때문에 낮은 H 와 PIC 값을 가짐을 알 수 있다. 또한 <그림 3>과 <그림 4>의 경우도 allele의 수는 2개로 같으나 allele 도수가 <그림 4>의 경우 균일하지 않으므로 <그림 3>에 비해 낮은 H 와 PIC 값을 가짐을 알 수 있다. 따라서 <그림 1>과 같이 다양한 종류의 allele을 가지고, allele 도수가 비슷한 마커가 개체 식별에 유용하다고 할 수 있다. 3절에서 선택된 13개의 DNA 마커들 중에서 마커의 정보력 H 값과 PIC 값을 구해 H 값이 높은 마커 순으로 정렬하였다. 그 결과는 다음 <표1>와 같다.

<표1> DNA 마커의 Allele의 도수확률과 H, PIC 값

DNA 마커	Alleles amplified									H	PIC
	1	2	3	4	5	6	7	8	9		
BMS941	0.016	0.032	0.161	0.274	0.113	0.081	0.161	0.097	0.065	0.839	0.820
BMS1242	0.109	0.109	0.172	0.188	0.094	0.25	0.078			0.834	0.815
TGLA44	0.097	0.129	0.21	0.21	0.161	0.177	0.016			0.828	0.805
HUJ223	0.076	0.121	0.273	0.106	0.227	0.076	0.121			0.822	0.804
BMS2519	0.242	0.081	0.177	0.242	0.161	0.081	0.016			0.812	0.786
IDVGA-2	0.129	0.048	0.032	0.194	0.065	0.194	0.306	0.032		0.806	0.780
EL03	0.167	0.167	0.167	0.303	0.152	0.045				0.799	0.771
BM3026	0.145	0.242	0.29	0.177	0.145					0.784	0.758
ILSTS035	0.086	0.414	0.086	0.034	0.034	0.069	0.086	0.138	0.052	0.778	0.758
BM3507	0.081	0.097	0.339	0.048	0.242	0.194				0.771	0.751
BMS1675	0.061	0.394	0.212	0.121	0.076	0.076	0.015	0.045		0.768	0.740
HW-YU2-CA03	0.183	0.183	0.017	0.333	0.233	0.05				0.765	0.729
HW-YU2-CA-95	0.207	0.138	0.069	0.31	0.276					0.761	0.746

BMS941은 H값이 0.8388, PIC값이 0.8202로 나타났다. 이와 같이 H값이 높고 다양한 종류의 allele을 가질 때 개체식별이 유용하다고 할 수 있다.

4. DNA marker 검정력 평가

개체식별에 유용하다고 선택된 DNA 마커들의 검정력을 평가하기 위해서 MP(Match Probability), R(Relatedness coefficient)을 사용한다(MP: Weir, 1996, R: Queller and Goodnight, 1989).

4.1 Match probability

MP는 독립인 두 동물 사이에 동일한 allele을 가질 확률 값으로, 유전자형의 도수들은 독립성이 보장되어야 함을 전제조건으로 하며, 다음과 같이 정의된다(Weir, 1996).

$$MP = \prod_{k=1}^m \left(\sum_{i=1}^{n_k} (p_{k_i}^2) + \sum_{i=1}^{n_k} \sum_{j=i+1}^{n_k} (2p_{k_i} p_{k_j}) \right) \quad (3)$$

where m = loci의 수

n_k = locus k의 allele의 수

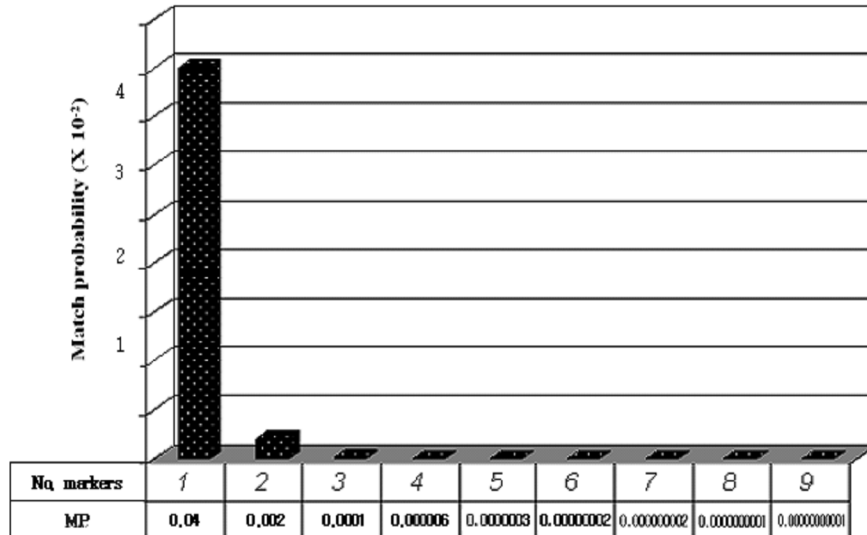
p_{k_i} = locus k의 i allele의 추정된 도수 비율

p_{k_j} = locus k의 j allele의 추정된 도수 비율

국가 후대 검정우 26-29차 36가계 305두 중 조부가 다른 33두를 사용하여, 개체 식별에 유용한 DNA 마커 13개를 <표1>에서 선택하였다. 선택된 13개의 마커 중에서 H 값이 높고 Hardy-Weinberg equilibrium을 만족하는 마커를 기준으로 9개(실험상의 기술적 문제, 시간, 비용을 고려한 마커수의 제한조건)의 누적 마커 조합으로 하는 44개의 후보 조합을 가지고 DNA 마커 검정력을 평가해 보았다. DNA 마커 검정력을 평가하기 위해 사용된 마커 수에 따른 MP값을 구해본 결과 확률 값이 가장 낮은 조합은 다음 <표2>, <그림4>와 같이 나타났다.

<표2> DNA 마커수의 변화와 MP값 조합

사용된 마커수	DNA marker 조합	MP
1	BMS941	0.04
2	BMS941 BMS1242	0.002
3	BMS941 BMS1242 TGLA44	0.0001
4	BMS941 BMS1242 TGLA44 HUIJ223	0.000006
5	BMS941 BMS1242 TGLA44 HUIJ223 BMS2519	0.0000003
6	BMS941 BMS1242 TGLA44 HUIJ223 BMS2519 ILSTS035	0.00000002
7	BMS941 BMS1242 TGLA44 HUIJ223 BMS2519 ILSTS035 BMS1675	0.000000002
8	BMS941 BMS1242 TGLA44 HUIJ223 BMS2519 ILSTS035 BMS1675 BMB507	0.0000000001
9	BMS941 BMS1242 TGLA44 HUIJ223 BMS2519 ILSTS035 BMS1675 BMB507 CA-03	0.000000000001



<그림4> 독립인 두 동물 사이에 동일한 allele을 가질 확률값(MP) (확률 값 X10⁻²)

<표2>와 <그림4>에서 한 개의 DNA 마커(BMS941)를 사용하였을 때 독립인 두 동물 사이에 동일한 alleles을 가질 확률(MP)은 100분의 4로 나타났고, 두 개의 DNA 마커(BMS941, BMS1242)를 사용하였을 때는 1000분의 2, 다섯 개의 DNA 마커(BMS941, BMS1242, TGLA44, HUJ223, BMS2519)를 사용하였을 때는 천만 분의 3, 아홉 개의 DNA 마커(BMS941, BMS1242, TGLA44, HUJ223, BMS2519, ILSTS035, BMS1675, BM3507, CA-03)를 사용하였을 때는 천억 분의 1로 나타났다. 이와 같은 결과를 바탕으로 이를 현실에 적용할 경우, 대구 경북지역에 있는 한우가 200000두라 가정 할 때, 6개의 마커를 사용하면 충분히 개체 식별이 가능하다는 것을 알 수 있다. 단, 각 개체들은 독립성이 보장되어야 함을 전제 조건으로 한다.

4.2 Relatedness coefficient

R계수(Relatedness coefficient)는 관련된 두 동물 사이의 닮음의 정도를 측정하는 도구로 사용되어 왔으며, MP가 독립성이 유지된 모집단에서 적용 될 수 있는 이론적인 값인 반면 R계수는 개별 혹은 자연적인 동물들 사이에 관련된 경험적인 확률(empirical probability)을 나타낸다(Queller and Goodnight, 1989).

$$R(i, j) = \frac{v(i, j)}{w(i)}$$

$$v(i, j) = \sum_{k=1}^2 \sum_{p=1}^2 \sum_{l=1}^r \sum_{m=1}^{s_r} \begin{cases} 0 & \text{if } x_k \neq y_p \neq m \\ [1 - f(m, l)] & \text{if } x_k = y_p = m \end{cases}$$

$$w(i) = \sum_{k=1}^2 \sum_{l=1}^r \sum_{m=1}^{s_r} \begin{cases} 0 & \text{if } x_k \neq m \\ [1 - f(m, l)] & \text{if } x_k = m \end{cases} \quad (4)$$

r = 마커의 수

s_r = 마커 r 의 alleles의 수

m = 마커 r 의 각각의 alleles

x_k = i 번째 동물에서 나타나는 allele, k 번째 위치, l 번째 locus

y_p = j 번째 동물에서 나타나는 allele, p 번째 위치, l 번째 locus

$f(m, l)$ = l 번째 locus에 마커 m 의 allele 도수 비율

위의 R 계수는 독립성조건 만족과 관련 없는 경험적 확률을 나타내므로 MP보다 더욱 현실적인 자료에 근접한 확률 값으로 볼 수 있다.

4.3 동물 오분류 확률(Probability of animal incorrectly identified)

R의 값이 0.85보다 크게 나타나면 더 많은 마커의 사용이 고려되어야한다(Cunningham et al., 1999). 이에 동물 오분류 확률은 다음과 같은 공식으로 나타낸다.

$$\text{동물 오분류 확률} = \sum \frac{I_{A(R)}}{\text{전체 비교의 수}} \quad (5)$$

$$I_A = \begin{cases} 1 & R \in A \\ 0 & R \notin A \end{cases} \quad A = \{R | R > 0.85\}$$

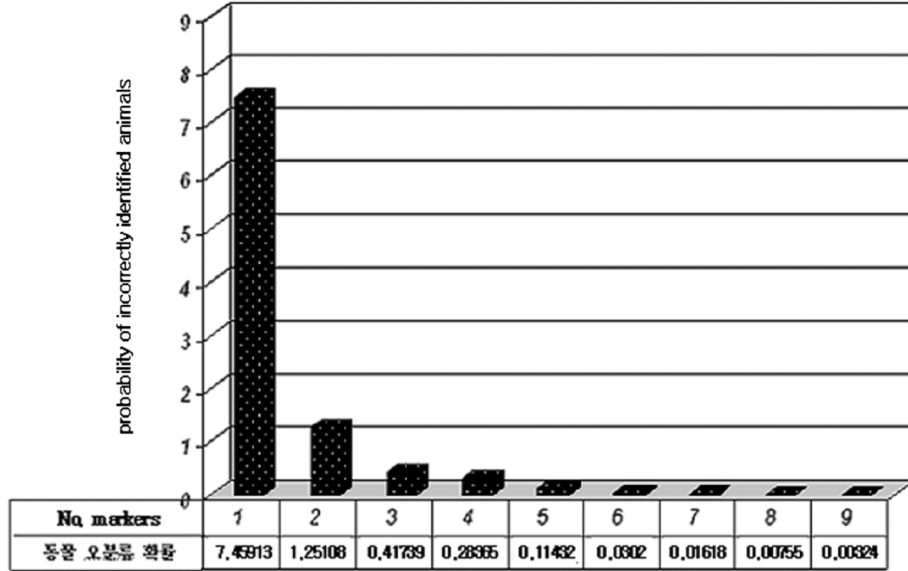
MP는 모집단이 <표1>과 같은 allele과 도수를 가지고 있고 임의로 두 개체의 alleles를 발생시켰을 때의 확률로 현실 데이터와는 차이가 나는 이론적인 값이므로 현실 적용에 다소 어려움이 따르는 것으로 보여 진다. 따라서 국가 후대 검정우 305두에서 선택된 조합을 R계수에 적용하여 동물 오분류 확률을 구해 DNA 마커 검정력 평가를 해보았다. 그 결과는 다음 <표3>, <그림5>와 같다.

<표3>, <그림5>에서 한 개의 DNA 마커(BMS2519)를 사용하였을 때 개체를 불확실하게 식별할 확률은 약 0.074, 두개의 DNA 마커(BMS2519, BMS1242)를 사용하였을 때 개체를 불확실하게 식별할 확률은 0.0125, 다섯 개의 DNA 마커(BMS2519, BMS1242, BMS941, BM3507, CA-95)를 사용하였을 때 개체를 불확실하게 식별할 확률은 0.0011, 아홉 개(실험상의 기술문제, 시간, 비용을 고려한 수)의 누적 DNA 마커(BMS2519, BMS1242, BMS941, BM3507, CA-95, TGLA44, ILSTS035, HUI223, CA-03)를 사용하였을 때 개체를 불확실하게 식별할 확률은 0.000032(즉, 100,000 마리 중 3.2마리)로 나타났다. 이와 같은 결과를 바탕으로 이를 현실에 적용할 경우, 대구 경북지역에 있는 한우가 200000두 라고 본다면 9개의 마커로 당장 현실 개체식별에 적용하기는 다소 불충분한 것으로 보여 진다.

<표3> DNA 마커 조합과 R 계수의 오분류 확률

사용된 마커수	DNA marker 조합	동물 오분류 확률
1	BMS2519	0.0745913
2	BMS2519 BMS1242	0.0125108
3	BMS2519 BMS1242 BMS941	0.0041739
4	BMS2519 BMS1242 BMS941 BM3507	0.0028365
5	BMS2519 BMS1242 BMS941 BM3507 CA-95	0.0011432
6	BMS2519 BMS1242 BMS941 BM3507 CA-95 TGLA44	0.00030198
7	BMS2519 BMS1242 BMS941 BM3507 CA-95 TGLA44 ILSTS035	0.00016178
8	BMS2519 BMS1242 BMS941 BM3507 CA-95 TGLA44 ILSTS035 HUI223	0.000075496
9	BMS2519 BMS1242 BMS941 BM3507 CA-95 TGLA44 ILSTS035 HUI223 CA-03	0.000032355

이와 같이 DNA마커 검정력을 평가해 본 결과 MP값이 천억 분의 일로 매우 낮은 확률 값을 가지지만 독립성이라는 조건을 만족해야 하므로 현실데이터에 적용하기에는 다소 어려움이 있는 것으로 보여진다. 따라서 독립성조건 만족과 관련이 없는 경험적 확률인 R계수에 적용한 동물 오분류 확률값을 이용하여 개체를 식별하는 것이 더 타당하다고 보여 지므로 BMS2519, BMS1242, BMS941, BM3507, CA-95, TGLA44, ILSTS035, HUI223, CA-03 DNA마커 조합이 개체 식별에 가장 유용하다고 할 수 있다.



<그림5> DNA 마커조합에 의한 R 계수의 오분류 확률 값

5. 결론 및 토의

국가 후대 검정우 26-29차 36가계 305두의 염색체에서 155개의 DNA 마커를 사용하여 개체식별에 유용한 DNA 마커(TGLA44, BMS2519, IDVGA-2, BMS1242, EL03, BM3026, ILSTS035, BMS941, HUI223, BM3507, BMS1675, HW-YU2-CA03, HW-YU2-CA95)를 찾았고, DNA마커 검정력을 평가 해 본 결과 독립성조건 만족과 관련이 없는 경험적 확률인 R계수에 적용한 동물 오분류 확률값이 현실데이터 적용에 더 유리하므로 개체 식별에 가장 유리한 DNA마커 조합은 BMS2519, BMS1242, BMS941, BM3507, CA-95, TGLA44, ILSTS035, HUI223, CA-03으로 밝혀졌다. 이에 따른 동물 오분류 확률은 대략적으로 100,000마리 중 3마리 정도로 나타났다. 한편, R 값에 적용한 동물 오분류 확률의 약점은 전체 비교 수 즉, 개체의 수에 영향을 받므로 개체수가 305두인 경우 전체 비교 횟수는 $305^2 - 305 = 92720$ 으로 고정되므로 확률 값을($\frac{1}{92720}$) 0.00001이하로 낮추는데 한계가 있었다. 따라서 이 실험의 결과를 현실에 적용할 경우, 대구 경북지역에 있는 한우가 200000두 라고 본다면 실험개체수가 적어도 500두($\frac{1}{249500} = 0.000004$) 정도는 되어야 현실 적용이 가능하다는 새로운 결론도 얻어졌다.

참고 문헌

1. Arana, A., Soret, B., Lasa, I and Alfonso L. (2002) Meat traceability using DNA markers : application to the beef industry. *Meat Science*. 61, 367-373.
2. Cunningham, E., Meghen, C., Scott, C., Bradley, D., MacHugh, D., Loftus, R. (1999) *DNA Traceability techniques for the meat industry*. In F. Toldra and Troy D.(Eds.) New developments in guaranteeing the optimal sensory quality of meat. 1-7 *Valencia*(Spain).
3. Glowatzki-Mullis, M., Gaillard, C., Wigger, G. and Fries, R.(1995) Microsatellite-based parentage control in cattle. *Anim. Genome*. 26, 7-12.
4. Lim, H., Min, H., Moon W., Lee, J., Kim, J. Cho, I., Lee, H., Lee, Y., Lee, J., and Jeon, J. (2005) Analysis and selection of microsatellites markers for individual traceability system in Hanwoo. *J. Anim. Sci & Technol*. 47(4), 491-500.
5. Ott. Jurg. (1999) *Analysis of human genetic linkage*, 3rd ed. The Johns Hopkins university press.
6. Peelman, L., Mortiaux, F., Zeveren, A., Dansercoer, A., Mommens, G., Coopman, F., Bouquet, Y., Burny, A., Renaville, R and portetelle, D. (1998) Evaluation of the genetic variability of 23 bovine microsatellite markers in Four Belgian cattle breeds. *Anim. Genet*. 29, 161-167.
7. Queller, D. C., & Goodnigh, K. F. (1989) Estimating relatedness using genetic markers. *Evolution*. 43(2), 258-275
8. Weir, B. (1996) *Genetic data analysis II. Methods for discrete population genetic data*. Sunderland, Massachusetts: Sinauer Associates, Inc. Publishers.
9. Weiss, K. (1993) *Genetic variation and human disease. principles and evolutionary approaches*. Cambridge, Great Britain: Cambridge University Press.
10. Yeo, J., Rhee, S., Choi, Y. Kwon, J. and Lee, J. (2006) Hanwoo(Korean Cattle) Traceability using DNA markers, *The Korean Communications in Statistics* Vol.13 No. 3, 2006, pp.733-743

[2007년 5월 접수, 2007년 5월 채택]