



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2009-0038737
(43) 공개일자 2009년04월21일

(51) Int. Cl.

C12Q 1/68 (2006.01)

(21) 출원번호 10-2007-0104192

(22) 출원일자 2007년10월16일

심사청구일자 2007년10월16일

(71) 출원인

경상대학교산학협력단

경상남도 진주시 가좌동 900

(72) 발명자

전진태

서울 서초구 방배2동 436-13번지 401호

김재환

제주 남제주군 성산읍 신흥리 812-1번지

(뒷면에 계속)

(74) 대리인

김순용

전체 청구항 수 : 총 16 항

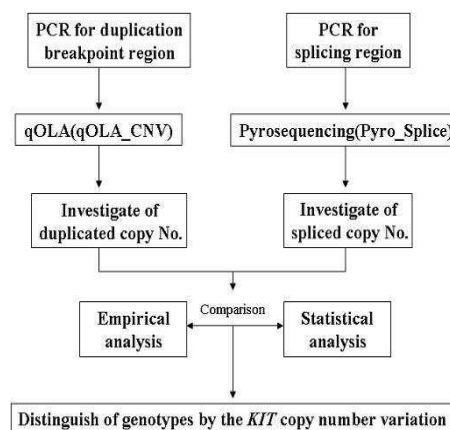
(54) 올리고뉴클레오타이드 연결 분석법을 이용한 케이아이티유전자 복제수 변이의 정량 분석방법

(57) 요약

본 발명은 올리고뉴클레오타이드 연결 분석법을 이용한 *KIT* 유전자 복제수 변이의 정량 분석방법에 관한 것이다. 보다 구체적으로 본 발명은 올리고뉴클레오타이드 연결 분석법을 이용한 *KIT* 유전자 중복의 복제수 변이 분석방법, *KIT* 유전자 중복의 복제수 변이는 상기 올리고뉴클레오타이드 연결 분석법에 의해 측정하고, *KIT* 유전자의 스플라이싱 변이는 종래의 파이로시퀀싱 법에 의해 측정하는 것을 특징으로 하는 *KIT* 유전자 복제수 변이의 정량방법, 상기 방법에 의해 측정된 *KIT* 유전자 복제수 변이에 따라 유전자형을 분류하는 것을 특징으로 하는 돼지 모색 관련 *KIT* 유전자의 유전자형 분석방법에 관한 것이다.

본 발명의 방법은 종래의 방법에 비해 보다 더 정확하고 정밀하게 *KIT* 유전자 복제수 변이의 정량할 수 있으므로, 본 발명의 방법은 *KIT* 유전자 중복 상태의 탐지와 유전자형 분석에 매우 유용하다.

대표도 - 도1



(72) 발명자

임현태

울산 중구 약사동 800-33번지

서보영

경남 진주시 옥봉동 670-2번지

이상호

경남 산청군 신안면 외송리 239번지

이재봉

경남 함양군 안의면 월림리 720-1번지

정은지

경남 사천시 사천읍 평화리 2-22번지 평화맨션 201호

유채경

경남 통영시 도천동 만복비치맨션 1차 1003호

이 발명을 지원한 국가연구개발사업

과제고유번호 20050301034467

부처명 농촌진흥청

연구사업명 바이오그린21사업

연구과제명 돼지 유전체 해독을 통한 유용유전자 발굴 및 활용기술개발

주관기관 축산과학원

연구기간 2007년 01월 01일~2007년 12월 31일

특허청구의 범위

청구항 1

(a) 분석하고자 하는 DNA 시료를 *LIMC1/LIME1* 지역에 특이적인 프라이머를 이용하여 PCR 증폭하는 단계;

(b) 상기 (a) 단계의 PCR 증폭산물을 단백질을 분해효소로 처리하는 단계; 및

(c) 상기 (b) 단계에서 단백질을 분해효소로 처리한 산물을 형광물질을 결합시킨 프라이머로 증폭시키고, *KIT* 좌위와 연관된 중복중지점에서 발생하는 유전자 중복(C/G)의 복제수 변이(copy number variation)를 측정하는 단계를 포함하는 폐지 모색 관련 *KIT* 유전자 중복의 복제수 변이 분석 방법.

청구항 2

제1항에 있어서, 상기 (a) 단계의 DNA 시료는 폐지의 혈액 또는 모근으로부터 분리된 것임을 특징으로 하는 방법.

청구항 3

제1항에 있어서, 상기 (a) 단계의 프라이머가 서열번호 1 및 서열번호 2로 표시되는 염기서열을 갖는 프라이머 세트 또는 서열번호 1 및 서열번호 3으로 표시되는 염기서열을 갖는 프라이머 세트인 것을 특징으로 하는 방법.

청구항 4

제1항에 있어서, 상기 (a) 단계에서 PCR 증폭은 94℃에서 20초, 58℃에서 20초, 및 72℃에서 30초를 27회 또는 25회 반복하는 것을 특징으로 하는 방법.

청구항 5

제1항에 있어서, 상기 (b) 단계의 단백질을 분해효소가 프로테아제 케이(proteinase K), 프로테아제 에이(proteinase A), 프로테아제 비(proteinase B), 또는 프로테아제 씨(proteinase C)임을 특징으로 하는 방법.

청구항 6

제1항에 있어서, 상기 (c) 단계의 프라이머는 서열번호 4 및 서열번호 5로 표시되는 염기서열을 갖는 프라이머 세트 또는 서열번호 4 및 서열번호 6으로 표시되는 염기서열을 갖는 프라이머 세트임을 특징으로 하는 방법.

청구항 7

제1항에 있어서, 상기 (c) 단계에서 증폭은 94℃에서 30초와 50℃에서 1분 30초를 10회 반복하는 것을 특징으로 하는 방법.

청구항 8

(a) 분석하고자 하는 DNA 시료를 *KIT* 유전자의 스플라이싱 변이 지역에 특이적인 프라이머를 이용하여 PCR 증폭하는 단계;

(b) 상기 (a) 단계의 PCR 산물을 대상으로 파이로시퀀싱 분석을 수행하여 *KIT* 좌위의 스플라이싱 돌연변이의 복제수 변이(copy number variation)를 측정하는 단계; 및

(c) 상기 제1항의 방법에 따라 측정된 *KIT* 좌위와 연관된 중복중지점에서 발생하는 유전자 중복(C/G)의 복제수 변이와 상기 (b) 단계에서 측정된 스플라이싱 돌연변이의 복제수 변이를 상호 비교하여 조합하는 단계를 포함하는 폐지 모색 관련 *KIT* 유전자의 복제수 변이 정량방법.

청구항 9

제8항에 있어서, 상기 (a) 단계의 *KIT* 유전자의 스플라이싱 변이 지역이 엑손 17과 인트론 17 사이의 지역임을 특징으로 하는 방법.

청구항 10

제8항에 있어서, 상기 (a) 단계의 프라이머는 서열번호 9 및 서열번호 10으로 표시되는 염기서열을 갖는 것임을 특징으로 하는 방법.

청구항 11

제8항에 있어서, 상기 (b) 단계의 스플라이싱 돌연변이가 *KIT* 유전자의 인트론 17의 첫번째 염기가 구아닌(G)에서 아데닌(A)으로 치환된 것임을 특징으로 하는 방법.

청구항 12

제8항의 방법에 의해 정량된 *KIT* 유전자의 복제수 변이에 의해 유전자형을 분리하는 것을 특징으로 하는 돼지 모색 관련 *KIT* 유전자의 유전자형 분석 방법.

청구항 13

제12항에 있어서, 상기 유전자형 분리는 관측된 특징들의 그룹핑 (grouping)을 바탕으로 한 좌표 분석의 분류와 근접 중심 소팅 (nearest centroid sorting) 방법을 이용하여 통계분석을 수행하는 것을 특징으로 하는 방법.

청구항 14

서열번호 4로 표시되는 염기서열을 갖는 프라이머.

청구항 15

서열번호 5로 표시되는 염기서열을 갖는 프라이머.

청구항 16

서열번호 6으로 표시되는 염기서열을 갖는 프라이머.

명세서

발명의 상세한 설명

기술 분야

<1> 본 발명은 올리고뉴클레오타이드 연결 분석법을 이용한 *KIT* 유전자 복제수 변이의 정량 분석방법에 관한 것이다.

배경 기술

<2> 돼지의 품종구별은 일반적으로 외형적 특징인 모색이나 체형 등의 표현형을 토대로 이루어진다. 하지만 표현형으로 구별되는 돼지 품종이라 할지라도 도축 후 육의 형태로 전환되면 돼지의 품종식별은 거의 불가능하게 된다. 또한 육질개선 등을 위한 3원 교잡종 등의 F2를 생산하는 현 육종체계에서 교잡에 사용된 종돈의 신뢰성 등 종돈으로써 고유한 특성의 보존이 문제시 되고 있다. 이러한 문제점들의 발생을 제거하기 위하여 돼지의 품종을 명확히 구분할 수 있는 DNA 마커의 선정과 활용이 필요하다고 사료된다. 현재 돼지에서 나타나는 품종 특이적인 DNA 마커를 개발하기 위하여 게놈 DNA와 미토콘드리아 DNA (mtDNA)내에 존재하는 여러 유전자들을 대상으로 활발하게 이루어지고 있다.

<3> 이 중에서 게놈 DNA 상에 위치하는 모색관련 유전자 중 *KIT* (mast/stem cell growth factor receptor)는 돼지의 우성백색과 유색을 조절하는 유전자로 돼지의 품종을 구별하는데 있어서 매우 중요한 DNA 마커로

사용되고 있다. 상기 *Dominant white/KIT* 좌위는 SSC8 (Johansson-Moller et al., *Mamm Genome*, 7:822-8, 1996; Hirooka et al., *J Hered*, 93:1, 2002) 상에 위치한다고 보고된 바 있다. 돼지의 백색 표현형은 두 개의 *KIT* 돌연변이(mutation)에 의해 결정된다: 유전자 중복은 불완전한 우성 표현형과 관련이 있으며, 스플라이싱 돌연변이는 완전한 우성 대립유전자 (allele)를 유도한다 (Johansson-Moller et al., *Mamm Genome*, 7:822-8, 1996; Marklund, S. et al., *Genome Res.*, 8:826-833, 1998). *KIT* 좌위에서 유색 표현형 (Color phenotype)의 열성 *i* allele, 패치 표현형 (Patch phenotype)의 I^P allele, 백색 표현형 (White phenotype)의 우성 *I* allele 그리고 벨트 표현형 (Belt phenotype)의 I^{Be} allele가 주된 대립유전자(allele)로 알려져 있다 (Giuffra, E. et al., *Mamm Genome*, 10:1132-113, 1999; Johansson M et al., *Genomics*, 14:965-96, 1992). Pielberg 등 (*Genome Res.*, 13:2171-217, 2003) 은 *I* allele의 다양성과 *I* allele를 I^1 , I^2 , I^3 그리고 I^L allele로 분류한 바 있다. 우성백색 유전자 *I*는 대립유전자 I^P 와 *i*에 대하여 완전 우성으로써 랜드레이스 (Landrace)와 라아지 화이트 (Large White)는 *I*를 하나 이상 소유하고 있고(I/I , I/I^P , 혹은 I/i .) 백색모를 나타내게 된다(Marklund, S. et al., *Genome Res.*, 8:826-833, 1998). 국내에서 가장 많이 사육되고 있는 착색모 품종인 듀록 (Duroc)은 i/i 이며, 이외에 버크셔 (Berkshire)와 햄프셔 (Hampshire)의 경우도 i/i 로 추정되고 있다(Marklund, S. et al., *Genome Res.*, 8:826-833, 1998).

<5> 현재 *KIT* 좌위의 분석 방법으로는 RFLP (Marklund, S. et al., *Genome Res.*, 8:826-833, 1998), 미니-시퀀싱(mini-sequencing)과 실시간 PCR (real-time PCR) (Pielberg, G. et al., *Genetics*, 160:305-3, 2002) 및 인베이터 (invader)와 파이로시퀀싱 분석법 (pyrosequencing assays) (Pielberg, G. et al., *Genome Res.*, 13:2171-217, 2003)이 활용된 바 있다. 이 중 파이로시퀀싱 분석법의 경우 실시간 PCR 또는 인베이터 기술에 비해 보다 더 정확한 결과를 나타내어 *KIT* 중복의 정량에 대해 가장 강력한 분석력을 가진 방법으로 사료되고 있다. 그러나 파이로시퀀싱 분석법은 단 하나의 복제에 의해서도 그 결과가 달라져 복제 수의 증가에 따라 점차적으로 유전자형 (genotype)을 정확하게 분류하기가 어려워 진다는 단점이 있다. 이것은 중복중지점이 점차 작아지는 징후가 증가하는 것과 연관이 있다 (Pielberg, G. et al., *Genome Res.*, 13:2171-217, 2003). 따라서 파이로시퀀싱 분석법을 이용한 대략적인 중복 비율의 탐지는 부모의 유전자형을 포함한 계통 정보를 모르는 샘플들에 대해서는 유전자형을 정확하게 판정하는 것이 어려운 단점이 있다.

<6> 이에 본 발명자들은 돼지 모색 관련 *KIT* 유전자의 복제수 변이를 보다 정확하게 측정할 수 있는 방법을 연구하던 중, 올리고뉴클레오타이드 연결 분석법(quantitative Oligonucleotide Ligation Assay)을 이용하여 높은 복제 수의 *KIT* 중복을 탐지할 수 있는 방법을 개발하고 상기 방법이 종래의 파이로시퀀싱 분석법에 비해 정확성 및 정밀성이 매우 우수함을 확인함으로써 본 발명을 완성하였다.

발명의 내용

해결 하고자하는 과제

<7> 따라서 본 발명의 목적은 올리고뉴클레오타이드 연결 분석법을 이용한 돼지 모색 관련 *KIT* 유전자 중복의 복제수변이 분석방법을 제공하는 것이다.

<8> 또한, 본 발명의 다른 목적은 *KIT* 유전자 중복의 복제수 변이는 올리고뉴클레오타이드 연결 분석법을 이용하여 측정하고, *KIT* 유전자의 스플라이싱 돌연변이의 복제수 변이는 파이로시퀀싱 분석법을 이용하여 측정하는 것을 특징으로 하는 *KIT* 유전자의 복제수변이 정량방법을 제공하는 것이다.

<9> 또한, 본 발명의 목적은 상기 방법에 의해 정량된 *KIT* 유전자의 복제수변이에 의해 유전자형을 분리하는 것을 특징으로 하는 돼지 모색 관련 *KIT* 유전자의 유전자형 분석방법을 제공하는 것이다.

<10> 또한, 본 발명의 목적은 상기 올리고뉴클레오타이드 연결 분석법을 이용한 돼지 모색 관련 *KIT* 유전자 중복의 복제수변이 분석방법에 유용한 프라이머 서열을 제공하는 것이다.

<11>

과제 해결수단

<12> 상기와 같은 본 발명의 목적을 달성하기 위하여, 본 발명은 올리고뉴클레오타이드 연결 분석법을 이용한 돼지 모색 관련 *KIT* 유전자 중복의 복제수변이 분석방법을 제공한다.

- <13> 또한, 본 발명은 *KIT* 유전자 중복의 복제수 변이는 올리고뉴클레오타이드 연결 분석법을 이용하여 측정하고, *KIT* 유전자의 스플라이싱 돌연변이의 복제수 변이는 파이로시퀀싱 분석법을 이용하여 측정하는 것을 특징으로 하는 *KIT* 유전자의 복제수변이 정량방법을 제공한다.
- <14> 또한, 본 발명은 상기 방법에 의해 정량된 *KIT* 유전자의 복제수변이에 의해 유전자형을 분리하는 것을 특징으로 하는 돼지 모색 관련 *KIT* 유전자의 유전자형 분석방법을 제공한다.
- <15> 또한, 본 발명은 상기 올리고뉴클레오타이드 연결 분석법을 이용한 돼지 모색 관련 *KIT* 유전자 중복의 복제수변이 분석방법에 유용한 프라이머 서열을 제공한다.
- <16>
- <17> 이하, 본 발명을 상세히 설명한다.
- <18>
- <19> 본 발명은 정량적 올리고뉴클레오타이드 연결 분석방법 (quantitative Oligonucleotide Ligation Assay, qOLA)을 이용하여 돼지 모색과 연관된 *KIT* 유전자의 중복을 분석하는 방법에 관한 것이다.
- <20> 올리고뉴클레오타이드 연결 분석방법은 DNA 리가아제 (DNA Ligase)를 이용하여 인접한 두 올리고뉴클레오타이드 프로브 (Probe)를 결합시키는 원리로 이용하는 방법이다.
- <21> 이를 위하여 본 발명에서는 현재 보고된 돼지의 모색과 밀접한 연관성이 있는 *KIT* 유전자의 중복중지점 (Duplication breakpoint)이 존재하는 *LIMC1/LIME1* 지역의 변이를 이용하였다.
- <22> 바람직하게는, 본 발명의 *KIT* 유전자 중복의 복제수변이 분석방법은 (a) 분석하고자 하는 DNA 시료를 *LIMC1/LIME1* 지역에 특이적인 프라이머를 이용하여 PCR 증폭하는 단계;
- <23> (b) 상기 (a) 단계의 PCR 증폭산물을 단백질을 분해효소로 처리하는 단계; 및
- <24> (c) 상기 (b) 단계에서 단백질을 분해효소로 처리한 산물을 형광물질을 결합시킨 프라이머로 증폭시키고 *KIT* 좌위와 연관된 중복중지점에서 발생하는 유전자 중복(C/G)의 복제수 변이 (copy number variation)를 측정하는 단계를 포함한다.
- <25> 상기 (a) 단계의 DNA 시료는 생물학적 시료로부터 분리될 수 있다. 상기 생물학적 시료로는 혈액 및 생물학적 기원의 기타 액상 샘플, 생검 표본, 모근, 조직 배양과 같은 고형 조직 샘플 또는 이로부터 유래된 세포 및 그 자손이 포함된다. 또한, 상기 샘플들은 시약처리, 가용화된 샘플 또는 배양세포, 세포 상등액 및 세포 용해물도 포함한다. 바람직하게는, 상기 DNA 시료는 돼지의 혈액 또는 모근으로부터 분리된 것일 수 있다.
- <26> 상술한 생물학적 시료로부터 DNA의 분리는 당업계에 공지된 방법에 따라 수행할 수 있다. 예를 들어, 혈액으로부터 게놈 DNA의 분리는 다음과 같은 방법으로 수행하였다. 혈액 10ml (0.5M EDTA 첨가)과 buffer A (1ℓ에 0.32M Sucrose, 1mM Tris-HCl pH7.5, 5mM MgCl₂, 1% Triton X-100 첨가)를 40ml 첨가한 후 여러 번 인버팅(inverting)하여 적혈구를 파괴시키고, 4℃에서 3,000rpm으로 10분간 스핀 다운(Spin down)시켜 백혈구를 침전시켰다. 상정액을 제거한 후 다시 buffer A를 15ml 첨가하여 인버팅(inverting) 후 4℃에서 3,000rpm으로 10분간 스핀 다운 (Spin down)시켜 백혈구를 침전시키고 상정액을 제거한 다음 buffer B (1ℓ에 400mM NaCl, 2mM EDTA pH8.0, 10mM Tris-HCl pH8.0 첨가) 4.5ml과 buffer C (5% SDS, 2mg/ml 프로테인아제 케이) 500ul를 첨가한 후 50℃의 수욕 (water bath)에서 오버나이트(overnight) 처리하였다. 상기 배양액에 6M NaCl 1.35ml을 첨가하여 강하게 인버팅(inverting) 한 후 3000rpm에서 15분간 스핀 다운(Spin down)시켜 상정액을 조심스럽게 새 튜브로 옮기고 2 부피(volume)의 100% 에탄올을 처리한 후 피펫-팁-훅 (pipette-tip-hook)를 이용하여 DNA를 건져냈다. 상온에서 알코올이 증발할 때까지 건조시킨 후 TE buffer (10mM Tris HCl, pH8.0; 2.5mM EDTA) 1ml에 녹여서 이용하였다.
- <27> 모근으로부터의 게놈 DNA 추출은 다음과 같은 방법으로 수행하였다. 모근 부착상태가 좋은 5~10가닥 내외의 모근을 적당한 길이로 자른 다음 5% 킬렉스(chelex) 100ul와 1mg/ml 프로테인아제 케이 1ul를 첨가한 후 55℃의 수욕 (water bath)에서 오버나이트 (overnight) 시켰다. 상기 샘플을 끓는 물에서 15분간 끓인 다음 상온에서 완전히 식힌 후 3000rpm에서 5분간 스핀 다운 (Spin down)시켜 PCR 반응에 상정액을 이용하였다.
- <28>
- <29> 상기 (a) 단계에서 *LIMC1/LIME1* 지역을 특이적으로 증폭시킬 수 있는 프라이머로는 *LIMC1/LIME1* 지역에

서만 특이적으로 나타나는 18 내지 25 bp의 연속된 각 폴리뉴클레오타이드를 포함할 수 있다.

- <30> 바람직하게는 *LIMC1/LIME1* 지역을 특이적으로 증폭시킬 수 있는 프라이머로는 서열번호 1 및 서열번호 2로 표시되는 염기서열을 갖는 프라이머 세트 또는 서열번호 1 및 서열번호 3으로 표시되는 염기서열을 갖는 프라이머 세트일 수 있다.
- <31> 상기 서열번호 2와 3은 극소량이 들어가서 기초 생산물을 형성한 후 실제적인 증폭은 서열번호 1과 TailR 프라이머 (서열번호 11)로 진행된다. 즉 서로 다른 염기를 갖는 서열번호 2와 서열번호 3은 genomic sequence로 만든 프라이머의 5' 말단에 공통적인 가상 sequence (TailR 포함)가 붙여져 있으며, 서열번호 1과 2 그리고 1과 3이 각각 반응하여 어느 정도 생산물을 증폭시킨 후 서열번호 1과 가상 프라이머인 TailR이 반응하여 소량으로 증폭된 각 생산물이 기하급수적으로 증폭되게 된다.
- <32> TailR primer : 5'-CGT TCG TAC GAG AAT CGC T-3' (서열번호 11)
- <33>
- <34> 상기 (a) 단계에서 PCR 증폭은 바람직하게는 94℃에서 20초, 58℃에서 20초, 그리고 72℃에서 30초를 27회 (모든 유래 DNA의 경우) 또는 25회 (혈액 유래 DNA의 경우) 수행할 수 있다.
- <35> 상기 (b) 단계에서 단백질분해효소는 *Taq* DNA 폴리머라제(Polymerase)의 활성을 제거하는 목적으로 처리될 수 있다. 상기 단백질분해효소로는 예를 들면, 이에 한정되지는 않으나 프로테아제 케이(proteinase K), 프로테아제 에이(proteinase A), 프로테아제 비(proteinase B), 프로테아제 씨(proteinase C) 등을 사용할 수 있다.
- <36> 상기 (c) 단계에서 형광물질이 결합된 프라이머는 *KIT* 유전자의 중복중지점을 기초로 하여 설계될 수 있다. 본 발명의 일 실시예에서는 공통 프라이머(BPT_Com)로서 중복중지점으로부터 -24th 뉴클레오타이드에서 -1st 자리까지의 염기를 포함한 서열번호 4로 표시되는 염기서열을 갖는 것과 중복(BPT1)과 정상적인 복제(BPT2)를 위한 각 역방향 프라이머(서열번호 5 및 서열번호 6)를 각각 설계하여 이를 사용하였다.
- <37> 또한, 상기 프라이머 중 BPT_Com의 5' 말단은 플루오레세인 (fluorescein)으로 표지하고, BPT1과 BPT2의 5' 말단에는 포스페이트 그룹 (phosphate group)을 첨가하였다. 또한, 두 개의 qOLA 산물을 구별하고 크기 (size) 구별 단계에서 사용되지 않은 프라이머의 피크로부터 떨어져있게 하기위해 BPT1과 BPT2의 3' 말단에 (dA)₁₀ 와 (dA)₁₅ 를 첨가하였다.
- <38> 상기 (c) 단계에서 증폭은 94℃에서 30초와 50℃에서 1분 30초를 10회 반복하여 수행하는 것이 바람직하다.
- <39>
- <40> 상기 (c) 단계에서 *KIT* 좌위와 연관된 중복중지점에서 발생하는 유전자 중복(C/G)의 복제수 변이는 리가아제 (ligase)를 이용하여 플루오레세인 (fluorescein)으로 표지한 BPT_Com 프라이머의 3' 말단에 포스페이트 그룹 (phosphate group)을 첨가한 BPT1과 BPT2 프라이머가 포스포디에스테르 결합 (phosphodiester bond)을 하도록 하여 결합된 부위를, 시퀀서 (sequencer, ABI3100)를 이용하여 중복된 유전자와 정상적인 유전자를 염기 길이 (base length)의 차이로 구별 후 중복된 유전자의 값을 전체 (중복+정상) 값으로 나누는 방법으로 측정할 수 있다.
- <41> 상기 본 발명에 따른 qOLA_CNV의 정밀성 및 정확도를 검증하기 위해서 기존의 파이로시퀀싱 분석법과 비교한 결과 표준곡선의 경우 qOLA_CNV와 Pyro_CNV 모두 좋은 선형성을 보였으나, 전반적인 Pyro_CNV에서 bias의 RMS (5.05) 및 SD (1.04)가 qOLA_CNV의 2.09와 0.45에 비하여 두 배 이상 높게 나타남을 확인할 수 있었다 (실시예 5 참조). 따라서, 본 발명의 방법이 종래의 파이로시퀀싱 분석법에 비해 보다 더 정밀성이 높음을 알 수 있었다.
- <42> 특히, 본 발명의 qOLA_CNV는 높은 복제 수 (> 4 per diploid genome)의 *KIT* 중복을 탐지하기위해 더 정확한 분석력을 가지고 있다.
- <43> 한편, 본 발명자들은 계통 정보가 없는 개체에 대한 높은 신뢰도를 위하여 선택적 스플라이싱 (splicing) 지역에서 발생하는 점 돌연변이 (point mutation)를 이용하였다. 즉, 중복중지점을 포함하고 있는 *LIMC1/LIME1* 지역에서 *KIT* 유전자 중복의 복제수 변이를 탐지하기 위해서 상술한 qOLA 분석법을 이용하였고,

KIT 유전자의 선택적 스플라이싱 (splicing) 지역에서 스플라이스 (splice) 형태의 복제수 변이를 탐지하기 위해서 기존에 보고 된 파이로시퀀싱 분석 방법 (Pielberg, G. et al., *Genetics.*, 160:305-3, 2002)을 이용하였다.

- <44> 따라서, 본 발명은 (a) 분석하고자 하는 핵산 시료를 *KIT* 유전자의 스플라이싱 변이 지역에 특이적인 프라이머를 이용하여 PCR 증폭하는 단계;
- <45> (b) 상기 (a) 단계의 PCR 산물을 대상으로 파이로시퀀싱 분석을 수행하여 *KIT* 좌위의 스플라이싱 돌연변이의 복제수 변이(copy number variation)를 측정하는 단계; 및
- <46> (c) 상기 제1항의 방법에 따라 측정된 *KIT* 좌위와 연관된 중복중지점에서 발생하는 유전자 중복(C/G)의 복제수 변이와 상기 (b) 단계에서 측정된 스플라이싱 돌연변이의 복제수 변이를 상호 비교하여 조합하는 단계를 포함하는 돼지 모색 관련 *KIT* 유전자의 복제수 변이 정량방법을 제공한다.
- <47> 상기 (a) 단계에서 *KIT* 유전자의 스플라이싱 변이 지역은 상기 유전자의 엑손17과 인트론17 사이를 말한다(Marklund, S. et al., *Genome Res.*, 8:826-833, 1998). 상기 지역을 특이적으로 증폭할 수 있는 프라이머로는 예를 들면, 이에 한정되지는 않으나, 서열번호 9 및 서열번호 10으로 표시되는 염기서열을 갖는 것일 수 있다.
- <48> 상기 (b) 단계에서 스플라이싱 돌연변이는 바람직하게는 *KIT* 유전자의 인트론 17의 첫번째 염기가 구아닌(G)에서 아데닌(A)으로 치환된 것을 말한다.
- <49> 상기 (b) 단계에서 스플라이싱 돌연변이의 복제수 변이는 프로브 프라이머를 이용하여 파이로시퀀싱 후 아데닌(A)과 구아닌(G)의 전체 면적에서 아데닌의 면적비율을 구하는 방법으로 측정할 수 있다.
- <50> 상기 (c) 단계에서는 본 발명의 qOLA 분석법에 의해 측정된 *KIT* 유전자 중복의 복제수 변이와 상기 파이로시퀀싱에 의해 측정된 *KIT* 유전자의 선택적 스플라이스 (splice) 형태의 복제수 변이를 조합하여 돼지 모색 관련 *KIT* 유전자의 복제수 변이를 정량한다.
- <51> 나아가, 본 발명은 상기 방법에 의해 정량된 *KIT* 유전자의 복제수 변이에 의해 유전자형을 분리하는 것을 특징으로 하는 돼지 모색 관련 *KIT* 유전자의 유전자형 분석 방법을 제공한다.
- <52> 상기 유전자형 분리는 관측된 특징들의 그룹핑 (grouping)을 바탕으로 한 좌표 분석의 분류와 근접 중심 소팅(nearest centroid sorting) 방법을 이용하여 통계분석에 의해 수행될 수 있다.
- <53> 본 발명의 일 실시예에서 qOLA_CNV 및 스플라이싱 형태인 *KIT* 복제의 양적 분석을 위한 파이로시퀀싱 분석(Pyro_Splice) 방법을 조합하여 계통 정보가 있는 145부 F₁의 유전자형 분석을 수행하였다(실시예 7 참조).
- <54> 그 결과, 본 발명의 방법에 의하면 *KIT* 유전자형 (genotype)을 정확하게 분류할 수 있음을 확인하였다(표 2 및 도 5참조). 이러한 결과는 qOLA_CNV와 Pyro_Splice의 조합을 바탕으로 한 유전자형 분리 방법을 높은 신뢰성과 계통 정보가 없는 임의의(random) 집단에도 활용할 수 있음을 보여준다.
- <55> 이에 본 발명자들은 본 발명의 다른 실시예에서 qOLA_CNV 및 Pyro_Splice 방법을 조합하여 임의로 선택된 100두의 상업돈 및 159두의 라아지 화이트 돈(Large White pigs)의 유전자형 분석을 수행하였다(실시예 8 참조). 그 결과, 본 발명의 방법은 이론적인 유전자형과 완전히 일치하는 진단결과를 나타냈다. 또한, 본 발명의 방법은 모근으로부터 유래된 DNA에 대해서도 민감하여 어떤 DNA에도 활용할 수 있음을 알 수 있었다(도 6 및 표 3 참조).
- <56> 나아가, 본 발명자들은 임의로 선택된 100두의 상업돈(CP)과 159두의 라아지 화이트돈(LW)에 대해 qOLA_CNV 또는 Pyro_CNV를 수행하고, 상기 두 방법을 비교하였다(실시예 9 참조). 그 결과, Pyro_CNV의 변이계수 값이 qOLA_CNV보다 두 배 이상 높게 나타났다(표 4 참조). 또한 159두의 LW의 유전자형을 분석한 결과 qOLA_CNV는 유전자형 분석에 있어서 명확한 결과를 보인 반면에, Pyro_CNV의 데이터는 많은 변이 (CV 범위: 7.3 - 15.3%)를 가지고 있는 것으로 나타났다. 따라서 qOLA_CNV가 Pyro_CNV보다 분석능이 더 우수하고 정확함을 알 수 있었다.

<57>

효 과

<58> 이상 살펴본 바와 같이, 본 발명의 방법은 종래의 방법에 비해 보다 더 정확하고 정밀하게 *KIT* 유전자 복제수 변이의 정량할 수 있으므로, 본 발명의 방법은 *KIT* 유전자 중복 상태의 탐지와 유전자형 분석에 매우 유용하다.

발명의 실시를 위한 구체적인 내용

<59> 이하, 본 발명을 실시예에 의하여 더욱 상세하게 설명한다. 단, 하기 실시예들은 본 발명을 예시하는 것으로 본 발명의 내용이 실시예에 의해 한정되는 것은 아니다.

<60> <실시예 1>

<61> 실험동물 및 DNA 시료의 추출

<62> 재래돈 (Korean native pig)과 랜드레이스 돈 (Landrace)을 상호교배 한 145두 F_1 의 *KIT* 대립인자 (allele)를 분리 분석하여 qOLA_CNV의 신뢰도 증명에 이용하였다. 또한, 듀록 (Duroc), 랜드레이스 (Landrace) 및 라아지 화이트 (Large White; LW) 종을 이용하여 3원교잡한 100두의 상업돈 (CP)을 무작위로 선발한 것과 159두의 LW를 대상으로 *KIT* CNV 실험을 하였으며 *KIT* 좌위에 대한 유전자형 (genotype) 분석을 하였다.

<63> F_1 과 CP의 혈액으로부터 레드 셀 라이시스-프로테아제 케이 방법 (red cell lysis-proteinase K, Miller, SA. et al., *Nucleic Acids Res*, 16:12, 1988)으로 게놈 DNA를 추출하였다. LW의 경우 LW 모근 DNA를 5% Chelex (Walsh, PS. et al., *Biotechniques*, 10:506-5, 1991)를 이용하여 준비하였다. 혈액에서 회수한 DNA는 TE 버퍼 (pH 8.0)에 용해하여 이용하였고, 모근 DNA는 그대로 PCR 주형가닥으로 이용하였다.

<64> <실시예 2>

<65> *KIT* 중복의 분석에 이용한 프라이머의 특이성 확인

<66> 본 발명에서는 *KIT* CNV의 분석을 위해 이전에 보고 된 PCR 프라이머를 사용하였다(Pielberg, G. et al., *Genome Res.*, 13:2171-217, 2003). 프라이머 서열은 반복되는 *L1MC1*과 *L1ME1*에 위치하는 *KIT* 중복중지점에서 선택되었다 (도 2의 a 참조). 정방향 프라이머 (KITBPF)(서열번호 1)는 *L1MC1* 서열과 80%의 서열 동일성을 보였으며, 정상적인 CNV를 위한 KIT1BPR (서열번호 2)과 중복된 CNV를 위한 KIT2BPR (서열번호 3) 역방향 프라이머는 *L1MC1*과 *L1ME1*에서 63.2%와 94.7%의 서열 동일성을 보였다. 이러한 특징은 결론적으로 PCR 증폭 산물이 부분적으로 다른 게놈 지역의 비 특이적인 것들을 포함하고 있을 것이라는 의문을 일으키게 한다. PCR 프라이머의 특이성에 대한 의문을 해결하기 위해 양적 분석을 하기 전에 벡터들로부터 형성된 패널에서 27개의 하이브리드 클론들을 체세포 하이브리드 패널 맵핑(Somatic cell hybrid panel mapping)법으로 분석하였다(Yerle, M. et al., *Cytogenet Cell Genet*, 73:194-2, 1996).

<67> 패널 안 교차 클론들의 PCR 반응은 전체 25 μ l 부피에 25ng 주형가닥 DNA, 1 μ l 반응 버퍼, 20pmol의 각 정방향 및 역방향 프라이머(KITBPF와 KIT1BPR; KITBPF와 KIT2BPR), 100 μ M dNTPs 및 5unit *Taq* DNA 폴리머라제 (GenetBio, Korea)를 첨가하였다. PCR 결과는 INRA (<http://www.toulouse.inra.fr/lgc/pig/pcr/pcr.htm>) 웹 페이지를 이용하여 분석하였다.

<68> 실험 결과, 정상적인 복제의 01000 11101 11000 11000 01000 10과 중복된 복제의 00000 11101 11010 11000 01000 10. 정상적인 복제와 중복된 복제의 위치는 사용된 primer set의 비 특이적 증폭을 나타내는 *KIT* 좌위 (assignment probability/correlation: 정상에 대해 0.8789/0.9250 그리고 중복된 것에 대해 0.8791/0.9250)가 존재하는 SSC8p11에서 선정되었다. 도 3에 나타낸 바와 같이 프라이머 세트는 선명하고 정확한 증폭을 하는 것을 확인할 수 있었다.

<69> <실시예 3>

<70> *KIT* 중복을 분석하기 위한 qOLA (qOLA_CNV)

<71> <3-1> qOLA_CNV를 위한 프라이머의 설계

<72> *KIT* 중복을 분석하기 위한 qOLA를 수행하기 위해 qOLA_CNV 프라이머를 설계하였다(도 2의 b). 공통 프라이머 (BPT_Com: 5'-Fluorescein- GGC TAC ATA CTG TAT GAT TCC AA -3', 서열번호 4)의 범주는 중복중지점으로부터 -24th 뉴클레오타이드에서 -1st 자리까지이다. 중복과 정상적인 복제를 위한 각 역방향 프라이머(BPT1:

5'-Phosphohate-GGG TCA TGG CTT GAA AAA GAA AAA AAA AAA-3', 서열번호 5 및 BPT2: 5'-Phosphate-CGA TAT GAC ATT CTG GAA ATA AAA AAA AAA AAA AA-3', 서열번호 6)는 중복중지점에서 설계되었다.

<73> BPT_Com의 5' 말단은 플루오레세인 (fluorescein)으로 표지하고, BPT1과 BPT2의 5' 말단에는 포스페이트 그룹 (phosphate group)을 첨가하였다. 또한, 두 개의 qOLA 산물을 구별하고 크기 (size) 구별 단계에서 사용되지 않은 프라이머의 피크로부터 떨어져있게 하기위해 BPT1과 BPT2의 3' 말단에 (dA)₁₀ 와 (dA)₁₅ 를 첨가하였다.

<74> <3-2> qOLA-CNV

<75> PCR은 20ng 게놈 DNA, 10pmol KITBPF (서열번호 1) 및 테일 프라이머 (tail primer), 0.1pmol의 테일 프라이머(tail primer) 결합 위치를 포함하는 각 역전사 프라이머 KIT1BPR (서열번호 2) 또는 KIT2BPR (서열번호 3), 200μM dNTPs, 1× 반응 버퍼 및 2.5unit Taq 폴리머라제 (GenetBio, Korea)를 이용하여 수행하였다. 열 조건은 94℃에서 20초, 58℃에서 20초, 그리고 72℃에서 30초를 27회 (모든 유래 DNA의 경우) 또는 25회 (혈액 유래 DNA의 경우)를 반복하였다.

<76> 상기에서 수득한 각 PCR 산물에 프로테이나제 케이(Proteinase K)를 최종 농도 10ug/μl가 되도록 첨가하고 55℃에서 40분 동안 반응시켰다.

<77> qOLA는 상기에서 프로테이나제 케이로 처리한 PCR 산물 1μl, 1× 반응 버퍼, 1.5unit Ampligase (EPICENTRE, USA), 1pmol BPT_Com (서열번호 4) 그리고 0.5pmol의 BPT1 프라이머 (서열번호 5) 또는 BPT2 프라이머 (서열번호 6)가 포함된 10μl의 혼합물을 이용하여 수행하였다. qOLA의 열 조건은 94℃에서 30초와 50℃에서 1분 30초를 10회 반복하였다. 그 다음 10μl의 탈이온 포름아미드 (deionize formamide)를 첨가하고 이를 ABI 프리즘 3100 제네틱 분석기(ABI Prism 3100 Genetic Analyzer, Applied Biosystems, USA)로 분석하였다. 피크에 상응하는 높이와 면적의 분석에는 진스캔 소프트웨어 버전 2.1(GeneScan software version 2.1, Applied Biosystems, USA)을 사용하였다. 데이터 수집과 기초적인 통계 계산은 마이크로소프트 엑셀 프로그램 (Microsoft, USA)을 이용하여 수행하였다.

<78> 상기 서열번호 1과 서열번호 2, 서열번호 1과 서열번호 3의 두쌍의 프라이머 세트와 증폭된 각각의 PCR 산물을 pGEM-T 벡터 (Stratagen, USA)를 이용하여 클론을 제조하였다. 클론의 증폭산물을 M13 정방향 프라이머 (서열번호 7) 및 역방향 프라이머(서열번호 8)를 이용하여 다시 증폭하였으며, 정제하여 증폭된 복제 및 정상적인 복제를 0%에서 100%까지 순차적으로 희석하였다. qOLA_CNV는 4반복의 PCR을 수행하였고, 피크의 높이와 면적에 대한 두개의 표준곡선(standard curve)을 추정하였다(도 4의 a 및 b). 높이와 면적 모두 표준곡선에서 상관관계수(correlation coefficient)가 0.999로서, 매우 좋은 선형성을 나타냈다.

<79> 상관관계수는 단지 표준곡선의 선형성만을 나타낸 것으로 정밀성 및 정확도를 비교하기 위해서 제곱평균 제곱근(root mean square, RMS) 및 표준편차(standard deviation, SD)를 비교하였다. 기준 비율의 치우친량을 의미하는 RMS 그리고 기준점 복제 시 변이량을 의미하는 SD는 예상한 것과 맞는 관측치 데이터를 상관관계수보다 더 잘 찾아내는 적절한 지표이다. TqOLA_CNV에서 피크 높이의 기대치가 기준값에 더 잘 맞고, 가장 낮은 변이를 보였다(표 1). 특히 전체에서 4개 이상의 KIT 복제를 가지는 개체를 정확히 유전자형 분석 (genotyping)하기 위한 정량 분석에는 표준곡선의 60%와 90% 사이의 구역에서 더 정확한 분석법이 필요하다. qOLA_CNV에서 피크 면적은 60%와 90% 사이의 구역에서 편향 (bias)의 RMS와 SD가 각각 2.21과 1.17로 나타났다. 반면에 qOLA_CNV에서 피크 높이는 같은 구역에서 0.86과 0.51로 나타났다.

<80> <실시예 4>

<81> KIT 중복을 분석하기 위한 파이로시퀀싱(pyrosequencing, Pyro_CNV)

<82> KIT CNV의 정량 분석을 위한 PCR과 파이로시퀀싱은 파이로마크 MD(PyroMark MD, Biotage, Sweden)을 이용하여 이전에 공지된 방법에 따라 수행하였다(Pielberg, G. et al., *Genome Res.*, 13:2171-217, 2003).

<83> <실시예 5>

<84> qOLA_CNV 및 파이로시퀀싱에 의한 표준곡선의 비교분석

<85> 실시예 3에서 수행한 qOLA_CNV와 실시예 4에서 수행한 Pyro_CNV를 비교하였다. qOLA_CNV에 이용된 순차적으로 희석한 것을 Pyro_CNV의 표준곡선 추정에 이용하였다. 그 결과, Pyro_CNV의 표준곡선은 좋은 선형성 (상관관계수 0.995)을 나타냈다(도 4의 c). 그러나, 전반적인 Pyro_CNV 편향(bias)의 RMS (5.05) 및 SD (1.04)

가 qOLA_CNV의 피크 높이 (2.09와 0.45)와의 비교에서 두 배 이상 높게 나타났다(표 1). 따라서, qOLA_CNV의 피크 높이를 이용한 돼지 *KIT*의 CNV 추정치 Pyro_CNV에 비해 시스템적 에러(error)와 변이가 보다 더 낮음을 알 수 있었다(도 4의 d).

표 1

표준곡선의 편향, 표준편차, RMS를 비교한 결과

Reference ^a (%)	Bias of qOLA_CNV (Height)	SD of qOLA_CNV (Height)	Bias of qOLA_CNV (Area)	SD of qOLA_CNV (Area)	Bias of Pyro_CNV	SD of Pyro_CNV
0	0	0	0	0	3.64	0.38
10	1.55	0.51	1.91	0.24	1.13	1.30
20	3.43	0.42	3.67	0.46	5.87	0.31
30	4.01	0.56	3.31	0.93	4.48	0.69
40	3.09	0.31	1.92	1.08	7.17	1.04
50	2.32	0.58	3.40	1.95	7.48	1.49
60	1.27	0.61	1.65	1.32	5.23	0.97
70	0.51	0.33	2.07	0.93	3.75	0.93
80	0.56	0.72	2.91	1.34	1.04	0.92
90	0.87	0.13	2.01	1.03	3.62	1.81
100	0	0	0	0	6.98	0.52
RMS1 ^b	2.09	0.45	2.16	1.03	5.05	1.04
RMS2 ^c	0.85	0.51	2.21	1.17	3.73	1.21

a: 도 4의 X축에서 복제된 카피의 기대치로서 기술된 표적값

b: 전체 RMS

c: 60-90% 사이 영역의 RMS

<실시예 6>

스플라이스 (Pyro_Splice) 형태의 *KIT* 복제에 대한 양적 분석

*KIT*21과 *KIT*35 PCR 프라이머 세트 (서열번호 9 및 서열번호 10)와 열 조건은 이전에 공지된 방법에 따라 수행하였다(Marklund, S. et al., *Genome Res* 1998, 8:826-8). 열 조건은 94℃에서 20초, 63℃에서 20초, 그리고 72℃에서 30초를 30회 반복하였다.

정방향 프라이머 *KIT*21: 서열번호 9

5'-GTA TTC ACA GAG ACT TGG CGG C-3'

역방향 프라이머 *KIT*35: 서열번호 10

5'-AAA CCT GCA AGG AAA ATC CTT CAC GG-3'

파이로시퀀싱은 Pielberg et al(*Genome Res.*, 13:2171-217, 2003)이 기술한대로 수행하였다. 파이로시퀀싱에 사용된 프로브 프라이머 Pielberg et al (*Genetics* 160: 305-311, January 2002)에 기술(5'-TAA TTA CIT GGT CAA AGG AAA C-3'(I는 이노신(Inosine)으로, 구아닌을 변환한 것이다.))한 바와 같으며, 스트렙타아비딘 비드 (Streptavidin bead, Streptavidin Sepharose High Performance, Amersham), 2X binding buffer 그리고 PCR 증폭 산물을 섞어 진공 조제 (Vacuum Prep)한 후 프로브 프라이머와 어닐링 버퍼(annealing buffer)의 혼합물에 떨어뜨려 80℃에서 2분간 변성(Denaturation) 시켰다.

<실시예 7>

qOLA CNV 및 Pyro_Splice 방법의 조합에 의한 유전자형 분석

KIT 대립유전자(allele)의 분리를 입증하기 위해, 실시예 3에서 수행한 qOLA_CNВ와 실시예 6에서 수행한 스플라이스(splice)된 형태인 *KIT* 복제의 양적 분석을 위한 파이로시퀀싱 분석(Pyro_Splice) 방법을 조합하여 재래돈 (Korean native pig)과 랜드레이스 돈 (Landrace)을 상호교배 한 145두 F₁의 유전자형 분석을 수행하였다.

qOLA_CNВ와 Pyro_Splice 방법의 조합에 의해 부모 세대 집단과 F₁ 집단의 유전자형을 분석할 수 있었다. 부모 세대 집단의 경우 수컷 8두와 암컷 11두인 19두의 재래돈(korean native)은 모두 블랙 표현형(Black phenotype)의 동형접합체 *i/i*을 보유하였으나, 수컷 8두와 암컷 9두의 17두의 랜드레이스(Landrace)는 화이트 표현형(White phenotype)의 3가지 다른 유전자형(genotype)인 *I*¹, *I*² 그리고 *i* 대립유전자(allele)의 조합을 보유하였다 (도 5의 a). F₁ 집단은 부모로부터 유전된 대립유전자의 조합인 이형접합체 *I*¹/*i*와 *I*²/*i* 그리고 동형접합체 *i/i*의 3가지 유전자형을 보유하였다 (도 5의 b).

유전자형 분리를 위해 관측된 특징들의 그룹핑 (grouping)을 바탕으로 한 좌표분석의 분류와 SAS 소프트웨어의 PROC FASTCLUS 절차(procedure)의 근접 중심 소팅(nearest centroid sorting) 방법을 이용하여 통계 분석을 수행하였다. 즉, 통계분석은 이론적인 유전자형의 스플라이스와 중복 비율을 12종류의 씨드(seed)로 정하고, 관측치를 유클리드 (Euclidean) 거리를 바탕으로 가장 가까이 있는 부류로 지정하여, 그 부류를 중심의 분류 표시와 같이 표시하였다.

유전자형 분석 결과 좌표분석 분류와 통계분석 분류 방법이 100% 일치하는 것으로 나타났으며, 이론적인 유전자형 비율 (표 2)과도 동일한 것으로 나타났다. F₁의 대립유전자들은 각 부모 세대의 랜드레이스 유전자형으로부터 정확하게 분리되었고, 이론적인 비율과 일치하는 판정 가능한 비율에서 클로스터 (cluster)되었다 (도 5의 c 및 d). 이러한 결과는 qOLA_CNВ와 Pyro_Splice의 조합을 바탕으로 한 유전자형 분리 방법을 높은 신뢰성과 계통 정보가 없는 임의의(random) 집단에도 활용할 수 있음을 보여준다.

표 2

스플라이싱 돌연변이 및 중복 특징을 이용한 *KIT* 좌위의 이론적 유전형

Genotypes ^a	Numbers of splice/normal	Ratios of splice (%)	Copy numbers of duplication/normal (total copy numbers)	Ratios of duplication (%)	Seed No. ^b
<i>i/i(I</i> ^{Be})	0/2	0	0/2(2)	0	1
<i>I</i> ^P / <i>i(I</i> ^{Be})	0/3	0	1/2(3)	33.3	2
<i>I</i> ^P / <i>I</i> ^P	0/4	0	2/2(4)	50	3
<i>I</i> ² / <i>I</i> ^P	1/4	20	3/2(5)	60	4
<i>I</i> ¹ / <i>I</i> ^P	1/3	25	2/2(4)	50	5
<i>I</i> ² / <i>i(I</i> ^{Be})	1/3	25	2/2(4)	50	5
<i>I</i> ¹ / <i>i(I</i> ^{Be})	1/2	33.3	1/2(3)	33.3	6
<i>I</i> ² / <i>I</i> ²	2/4	33.3	4/2(6)	66.7	7
<i>I</i> ¹ / <i>I</i> ²	2/3	40	3/2(5)	60	8
<i>I</i> ³ / <i>I</i> ^P	2/3	40	3/2(5)	60	8
<i>I</i> ¹ / <i>I</i> ¹	2/2	50	2/2(4)	50	9
<i>I</i> ³ / <i>i(I</i> ^{Be})	2/2	50	2/2(4)	50	9
<i>I</i> ² / <i>I</i> ³	3/3	50	4/2(6)	66.7	10
<i>I</i> ¹ / <i>I</i> ³	3/2	60	3/2(5)	60	11
<i>I</i> ³ / <i>I</i> ³	4/2	66.7	4/2(6)	66.7	12

a: I^1 대립유전자는 포함되지 않으며 유전자형은 스플라이싱 돌연변이의 비율에 의해 정렬되었음.

b: 통계적인 유전자형 지정에 사용된 중심 계급(class centroids)의 수치임.

<실시예 8>

qOLA_CNV 및 Pyro_Splice 방법의 조합에 의한 상업돈 및 라아지 화이트 돈의 분석

qOLA_CNV 및 Pyro_Splice 방법을 조합하여 임의로 선택된 100두의 상업돈 및 159두의 라아지 화이트 돈 (Large White pigs)의 유전자형 분석을 수행하였다. qOLA_CNV 및 Pyro_Splice는 실시예 4 및 실시예 6과 동일한 방법으로 수행하였다.

먼저, 100두의 상업돈 (CP)의 경우 종결 부돈 듀록 (Duroc)은 열성 동형접합체 i/i 로 이루어져 있으며, 모돈으로 이용한 F_1 은 표현형적으로 화이트 (White)이며, 동형접합체 또는 이형접합체 상태의 우성 I 대립유전자를 가지고 있을 것으로 추정된다. 따라서 CP는 적어도 듀록으로부터 유전된 하나의 i 대립유전자를 가질 것으로 예상된다. qOLA_CNV 및 Pyro_Splice 방법의 조합에 의해 관측된 특징을 4 종류로 그룹핑하였다: I^1/i , I^2/i , I^P/i 및 $i(I^{Be})/i$ (도 6의 b). 관측된 특징을 그룹핑한 것과 통계적 클러스터링 한 것이 완벽하게 일치하였다.

159두의 라아지 화이트 돈의 경우 CP에 비하여 LW에서의 CNV (3~6 복제)의 범주가 더 넓을 것이라 예상되었고, 이는 qOLA_CNV 분석의 연장 평가라 할 수 있다. 더욱이, LW DNA는 모근으로부터 준비하였기 때문에 qOLA_CNV의 민감성을 추정할 수 있다. 먼저, 좌표 분석 분류를 수행하였다; LW를 실험에 의해 확보한 데이터의 그룹화되는 특징을 바탕으로 11종의 유전자형을 포함하는 8 종류로 분류하였다 (도 6의 c).

그 다음에, 분류한 것을 기초로 SAS 소프트웨어의 PROC FASTCLUS 절차(procedure)의 근접 중심 소팅 (nearest centroid sorting) 방법을 이용하여 통계분석을 수행하였다. 그 결과, 두 유전자형 분석 방법 모두 두 개체 (IDs: T7183와 Y4410Q)를 제외하고 동일한 결과를 나타냈다. T7183개체는 qOLA_CNV와 Pyro_Splice가 63.4%와 62.9%이며, 점의 좌표가 I^1/I^3 표준 좌표 (60%와 60%)와 명확히 구분되었다. 그러나 좌표와 I^3/I^3 중심 (66.7%와 66.7%) 사이의 거리가 4.78로, 좌표와 I^1/I^3 중심 사이의 4.45보다 조금 더 길었다(표 3).

표 3

159두의 라아지 화이트 돈에 대한 관측치 및 통계치의 비교

	4 ^a	5 ^a	7 ^a	8 ^a	9 ^a	10 ^a	11 ^a
I^2/I^P	6 (8.77%)	-	-	-	-	-	-
I^1/I^P or $I^2/i(I^{Be})$	-	12 (7.55)	-	-	-	-	-
I^2/I^2	-	-	28 (14.47)	1 (0.63)	-	-	-
I^1/I^2 or I^3/I^P	-	-	-	58 (36.48)	-	-	-
I^1/I^1 or $I^3/i(I^{Be})$	-	-	-	-	51 (32.08)	-	-
I^2/I^3	-	-	-	-	-	5 (3.14)	-
I^1/I^3	-	-	-	-	-	-	2 (1.26)
I^3/I^3	-	-	-	-	-	-	1 (0.63)

* 관찰된 그룹핑 특징에 의해 지정된 유전자형은 종렬에 나타내었고 통계 분석을 위한 중심 계층(class centroids)의 수는 횡렬에 나타내었다.

*두 방법 간의 2개의 불일치된 결과는 이탤릭체 및 굵은 글씨체로 나타내었다.

a: 각 유전자형의 중심 계층(seed number)의 수

<실시예 9>

상업돈 및 라이지 화이트 돈의 분석에 있어서 qOLA_CNV 및 Pyro_CNV 의 비교

100두의 상업돈(CP)과 159두의 라이지 화이트돈(LW)에 대해 qOLA_CNV와 Pyro_CNV를 수행하고, 상기 두 방법을 비교하였다.

그 결과, 70두 I^1/i CP의 변이계수 (coefficient of variation, CV)는 qOLA_CNV의 경우 2.9%이며, Pyro_CNV의 경우 6.9%였다. Pyro_CNV의 CV 값이 qOLA_CNV보다 두 배 이상 높게 나타났다(표 4). 이는 두 방법에 의한 표준곡선(standard curve)의 비교에서 나타나는 것과 동일한 경향이다(실시예 5).

또한 159두의 LW를 FASTCLUS 절차를 이용하여 클러스터 씨드(cluster seed) 근처에서 선정된 관측치의 유전자형을 분석한 결과 qOLA_CNV는 유전자형 분석에 있어서 명확한 결과를 보인 반면에, Pyro_CNV의 데이터는 많은 변이 (CV 범위: 7.3 - 15.3%)를 가지고 있는 것으로 나타났다. 따라서 qOLA_CNV가 Pyro_CNV보다 분석능이 더 우수하고 정확함을 알 수 있었다.

표 4

상업돈 및 라이지 화이트돈의 유전자형 분석에 있어서 qOLA_CNV, Pyro_CNV 및 Pyro_Splice의 정확성 및 정밀성 비교

Breed or Cross (Type of DNA)	Estimated Genotype by qOLA	No. of pigs	qOLA (%)			Pyrosequencing (%)					
			Expected ^a	Observed ^{bd} in qOLA_CNV	CV ^c	Expected ^a	Observed ^{bd} in Pyro_CNV	CV ^c	Expected ^a	Observed ^{bd} in Pyro_Splice	CV ^c
상업돈 (혈액유래 DNA)	I^1/i	70	33.3	33.7	2.9	33.3	31.8	6.9	33.3	35.9	2.2
	I^2/i	23	50	49.4	1.8	50	44.7	7.2	25	27.7	2.6
	I^P/i	5	33.3	34.3	na ^g	33.3	29.7	na	0	0	na
	$i(I^{Be})/i$	3	0	0	na	0	1.8	na	0	0	na
라이지 화이트돈 (모근유래 DNA)	I^1/I^1 or $I^3/i(I^{Be})$	51	50	50.7	3.4	50	47.1	11.7	50	51.4	2.1
	I^1/I^2 or I^3/I^P	58	60	58.2	3.4	60	54.6	15.9	40	41.8	2.4
	I^1/I^P or $I^2/i(I^{Be})$	12	50	51.0	3.1	50	47.4	7.3	25	27.9	3.8
	I^2/I^2	24	66.7	65.7	3.5	66.7	61.1	11.3	33.3	34.3	6.2
	I^1/I^3	2	60	57.3	na	60	58.3	na	60	61.2	na
	I^2/I^P	6	60	59.2	na	60	52.3	na	20	22.8	na
	I^2/I^3	5	66.7	64.5	na	66.7	60.3	na	50	50.4	na
	I^3/I^3	1	66.7	63.4	na	66.7	58.7	na	66.7	62.9	na

a: 각 유전자형에 상응하는 이론적인 비율

- <126> b: 각 유전자형에 있어서 관찰된 비율의 평균값
- <127> c: 변이계수. 10개 이상의 유전형을 가진 개체를 포함하는 4개의 유전자형이 분석을 위해 선별됨.
- <128> d: *KIT* 중복에 대한 올리고뉴클레오타이드 결합 분석
- <129> e: *KIT* 중복에 대한 파이로시퀀싱 비율
- <130> f: *KT*의 스플라이스 돌어변이에 대한 파이로시퀀싱 비율
- <131> g: na는 분석되지 않음을 의미함

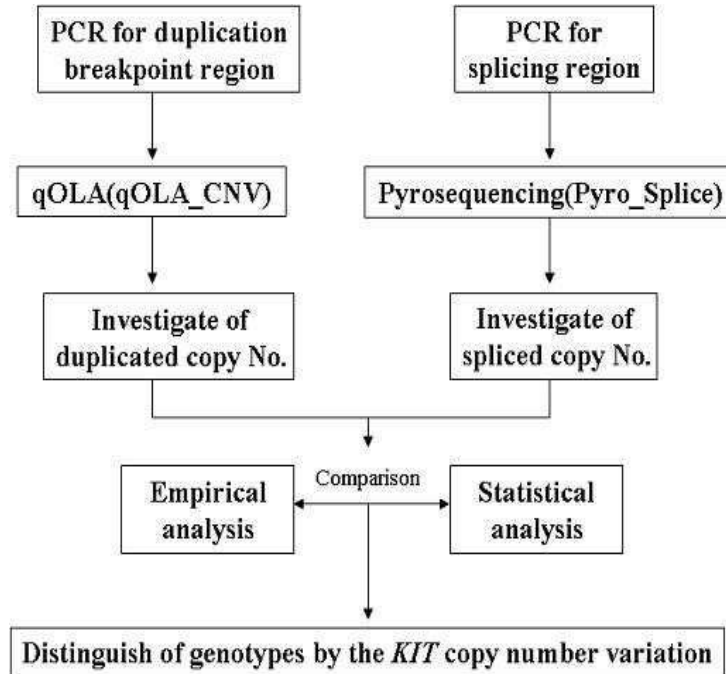
도면의 간단한 설명

- <132> 도 1은 본 발명에 따른 돼지 모색 관련 *KIT* 유전자의 복제수 변이 탐지방법 및 파이로시퀀싱 방법을 비교하여 도식화한 것이다.
- <133> 도 2의 돼지 *KIT* 좌위와 중복의 전체적인 모식도(a) 및 중복중지점 주변의 염기서열(b)을 나타낸 것이다.
- <134> 도 3은 체세포 하이브리드 패널을 이용한 PCR 분석 결과이다.
- <135> a: 서열번호 1과 서열번호 2 프라이머 세트의 특이성을 검증하기 위해 체세포 하이브리드 패널 결과이다.
- <136> b: 서열번호 1과 서열번호 3 프라이머 세트의 특이성을 검증하기 위해 체세포 하이브리드 패널 결과이다.
- <137> 각 레인의 수치는 체세포 하이브리드 패널에 이용한 클론 (총 27개) 중 PCR 증폭된 클론의 번호를 표기한 것이다.
- <138> 도 4는 qOLA_CNV와 Pyro_CNV의 표준곡선이다.
- <139> a, b, c의 X 축: 중복 복제수의 각 기대치들을 표기한 것이다
- <140> a와 b의 Y축: qOLA를 수행한 결과 값들으로써 a의 Y축은 각 기대치에 대한 qOLA_CNV 관측치 중 피크 높이의 실제 값을 표기한 것이며, b의 Y축은 각 기대치에 대한 qOLA 관측치 중 피크 면적의 실제 값을 표기한 것이다.
- <141> c의 Y축: 파이로시퀀싱을 수행한 결과 값들으로써 Pyro_CNV의 실제 값을 표기한 것이다.
- <142> d: 표1의 RMS1을 나타낸 것으로 X축은 |Bias| 값을 Y축은 SD 값을 각각 표기 한 것이다. 즉, X축은 실제적 절대치의 편차를 표기한 것이고 Y는 반복 내의 변이를 표기한 것이다.
- <143> 도 5는 *KIT* 복제 수 분석을 위한 qOLA_CNV와 스플라이싱 형태의 *KIT* 복제의 양적 분석을 위한 파이로시퀀싱 (Pyro_Splice)을 이용한 *KIT* 대립유전자의 분리 분석에 의한 유전자형을 분류한 결과이다.
- <144> a 및 b: qOLA 결과(qOLA_CNV)를 X축에 표기하고 파이로시퀀싱 결과(Pyro_Splice)를 Y축에 표기한 것이다. a는 부모돈들의 비율을 b는 자돈들의 비율을 각각 표기한 것이며, 각 유전자형으로 그룹핑이 된다는 것을 보여 준 것이다.
- <145> c 및 d: 자돈들이 각 부모돈들의 유전자형에 어떻게 분포가 되어 있는지를 보여주는 것으로써 X축은 부모의 유전자형을 표기한 것이며 Y축은 자돈들의 비율을 표기한 것이다. c는 자돈들의 qOLA 분포도를 d는 파이로시퀀싱 분포도를 각각 표기한 것이다. 부모돈들은 한쪽은 랜드레이스 (Landrace)이며 다른 한쪽은 듀록 (Duroc(i/i))이므로 c와 d의 X축에 고정된 듀록 유전자형은 표기 하지 않았다.
- <146> e: 각 유전자형의 qOLA와 파이로시퀀싱 수행시 실제로 어떻게 나타나는지를 보여 준 것이다.
- <147> 도 6은 qOLA_CNV와 Pyro_Splice 조합을 이용하여 상업돈 및 라아지 화이트 돈의 유전자형을 분석한 결과이다.
- <148> a: 통계분석에 이용되는 12종류의 씨드(seed)의 위치를 표기한 것이다.
- <149> b 및 c: 좌표 분석을 표기한 것으로 씨드 주위에 각 결과값들이 밀집되는 것을 보여 준 것이다. c의

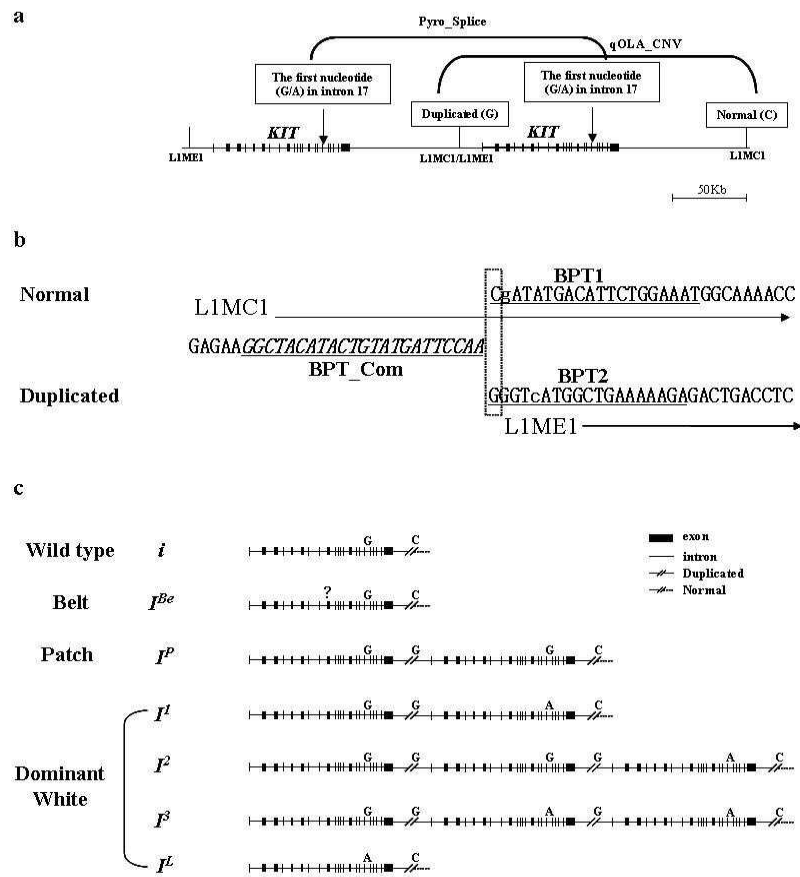
T7183과 Y4410Q는 표 3에서의 두개의 불일치 되는 것들(좌표분석 분류와 통계분석 분류 비교시)을 표기한 것이고, b는 상업돈들의 그룹이며 c는 라아지 화이트 그룹이다.

도면

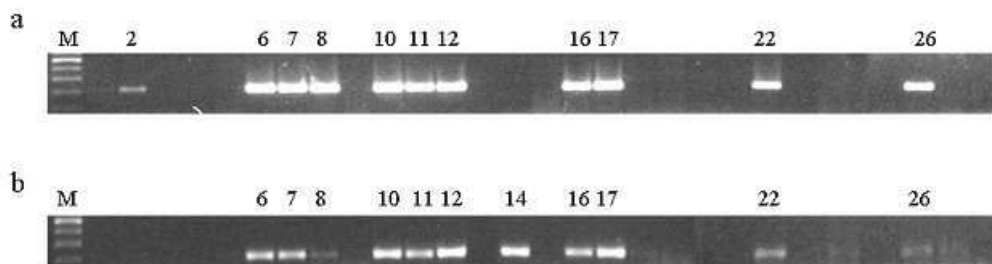
도면1



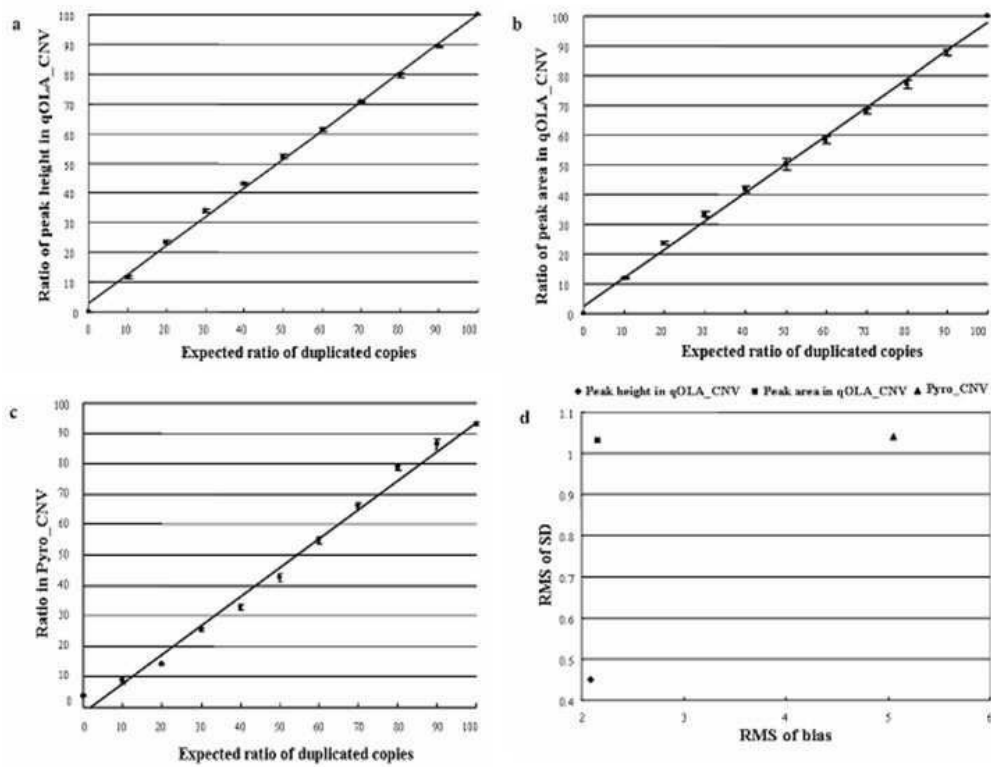
도면2



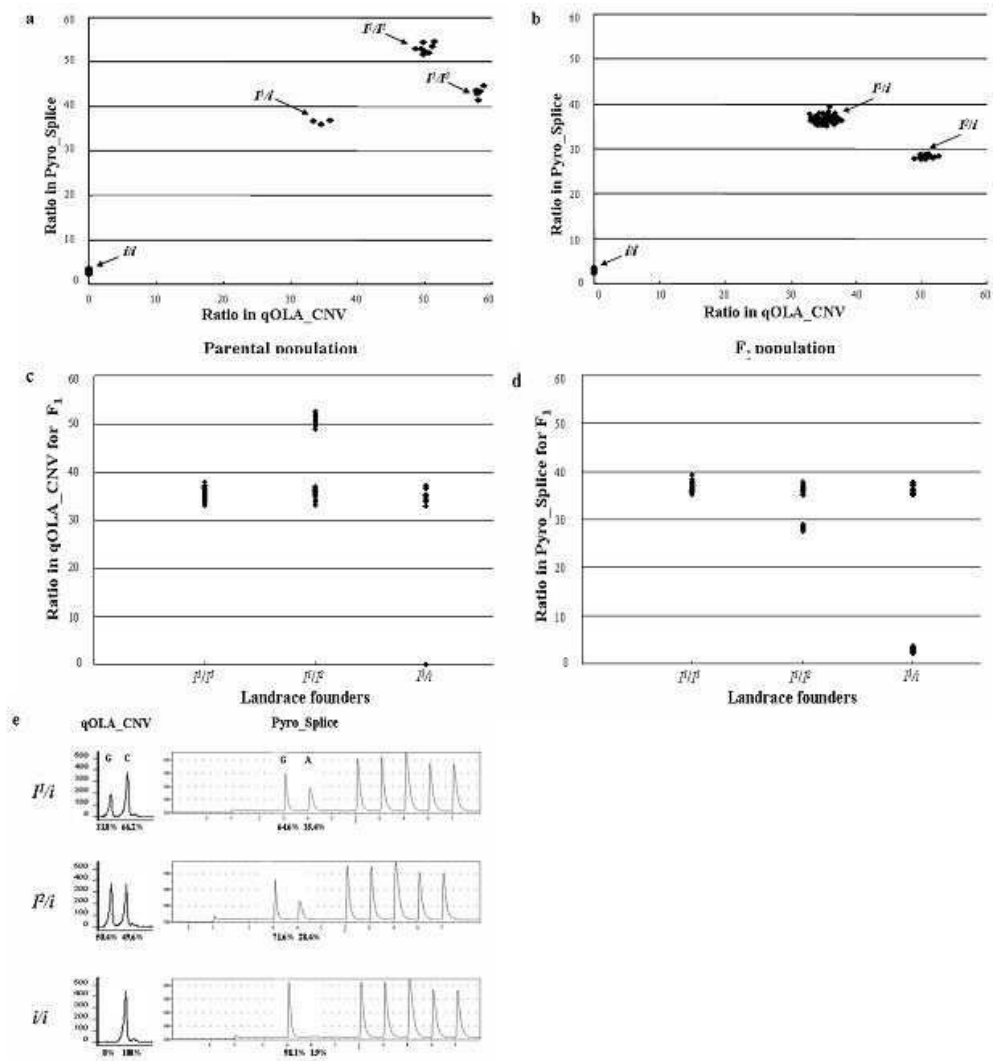
도면3



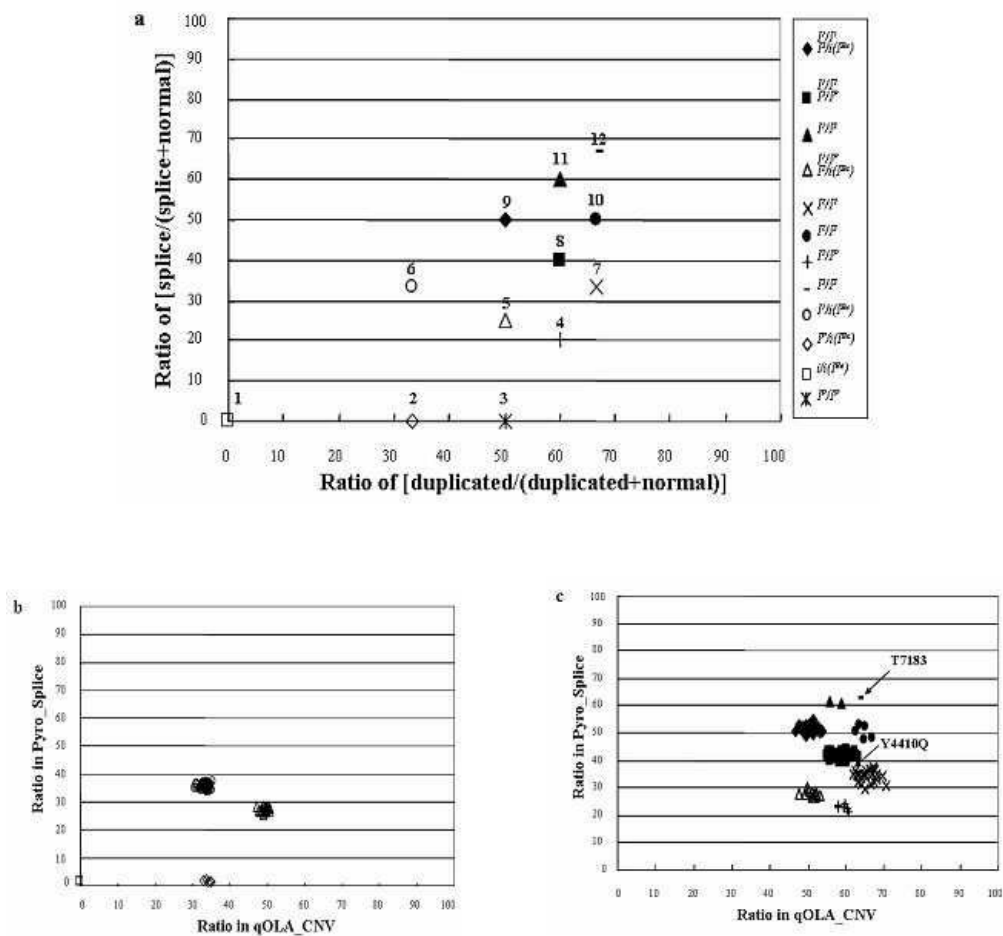
도면4



도면5



도면6



서열 목록

<110> INDUSTRY-ACADEMIC COOPERATION FOUNDATION GYEONGSANG NATIONAL UNIVERSITY
 <120> Method for quantifying and analyzing copy number variation in porcine KIT by an Oligonucleotide Ligation Assay

<160> 11

<170> Kopatent In 1.71

<210> 1

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> forward primer KITBPF

<400> 1
caactatgct acatccaggc 20

<210> 2
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Reverse primer KIT1BPR

<400> 2
gtaaccgttc gtacgagaat cgctgtggct cttttgaggt cag 43

<210> 3
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Reverse primer KIT2BPR

<400> 3
gtaaccgttc gtacgagaat cgctctgtct ccatggtttt gcc 43

<210> 4
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> primer BPT_Com

<400> 4
ggctacatac tgtatgattc caa 23

<210> 5
<211> 30
<212> DNA

<213> Artificial Sequence

<220>

<223> primer BPT1

<400> 5

gggtcatggc ttgaaaaaga aaaaaaaaaa

30

<210> 6

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> primer BPT2

<400> 6

cgatatgaca ttctggaaat aaaaaaaaaa aaaaa

35

<210> 7

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> M13 forward primer

<400> 7

cacgacgttg taaaacgac

19

<210> 8

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> M13 Reverse primer

<400> 8

ggataacaat ttcacacagg	20
<210> 9	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Forward primer KIT21	
<400> 9	
gtattcacag agacttggcg gc	22
<210> 10	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Reverse primer KIT35	
<400> 10	
aaacctgcaa ggaaatcct tcacgg	26
<210> 11	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> TailR primer	
<400> 11	
cgttcgtacg agaatcgct	19