

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2011-0021095 (43) 공개일자 2011년03월04일
(51) Int. Cl. <i>C12N 15/11</i> (2006.01) <i>C12Q 1/68</i> (2006.01) (21) 출원번호 10-2009-0078690 (22) 출원일자 2009년08월25일 심사청구일자 없음	(71) 출원인 울산대학교 산학협력단 울산 남구 무거2동 산29 (72) 발명자 이종극 서울특별시 강남구 도곡2동 개포한신아파트 4동 708호 김재정 서울특별시 동대문구 제기2동 현대아파트 103동 503호 (뒷면에 계속) (74) 대리인 특허법인태백	

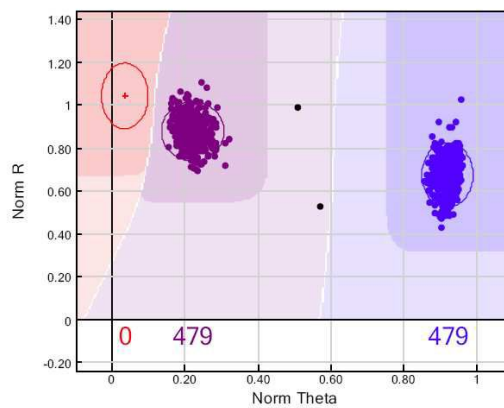
전체 청구항 수 : 총 6 항

(54) 성별 판별 및 / 또는 개체 식별용 SNP 마커

(57) 요약

본 발명은 단일염기변이를 갖는 SNP(single nucleotide polymorphism) 마커를 이용한 최적의 성별 판별 및/또는 개체 식별용 마커 세트와 이를 이용한 성별 판별 및/또는 개체 식별 방법에 관한 것으로, 상기 SNP 마커를 이용하면 종래 사용하던 STR 마커와 달리 낮은 돌연변이율, 분해된 DNA 시료에 유용한 100bp보다 짧은 앰플리콘 크기의 유용성 및 고효율 지노타이핑 기술 등의 장점이 있어 성별 판별 및 개체 식별에 매우 유용하다.

대표도 - 도5



(72) 발명자

이해인

경기도 구리시 교문동 766-1 한성아파트 110동 40
6호

한복기

서울특별시 종로구 평창동 72번지 롯데아파트 101
동 1104호

이 발명을 지원한 국가연구개발사업

과제고유번호 01-PJ10-PG6-01GN15-001 , 2008-S6-E-001

부처명 보건복지부 (한국보건산업진흥원) , 보건복지부 질병관리본부 (용역사업)

연구관리전문기관

연구사업명 보건의료연구개발사업 , 용역사업

연구과제명 유전체기법을 이용한 선천성 기형 및 유전질환의 원인 유전자 탐색과 질병예측법 개발 ,
인체유래 생물자원 활용화 연구 시범사업

기여율

주관기관 울산대학교 산학협력단

연구기간 2006년 11월 01일 ~ 2011년 10월 31일 , 2008년 08월 26일 ~ 2009년 03월 15일

특허청구의 범위

청구항 1

아멜로제닌(Amelogenin) 유전자(AMELX & AMELY)에서 발굴된 성별 판별용 SNP 마커 3개와 아연 핑거 단백질 유전자(ZFX & ZFY)에서 발굴된 성별 판별용 SNP 마커 3개를 포함한 총 6개로 이루어진 성별 판별용 SNP 마커.

청구항 2

청구항 1에 있어서, 상기 아멜로제닌 유전자(AMELX & AMELY)의 성별 판별용 SNP 마커는 NCBI GenBank에 등록된 AMELX 유전자 염기서열(GenBank Accession No: NC_000023) 정보에서 염기서열 #5139 (C/T), 염기서열 #5196 (C/G), 그리고 염기서열 #5232 (C/T)에 위치하는 것을 특징으로 하는 성별 판별용 SNP 마커.

청구항 3

청구항 1에 있어서, 상기 아연 핑거 단백질(ZFX & ZFY) 유전자의 성별 판별용 SNP 마커는 NCBI GenBank에 등록된 ZFX 유전자 염기서열(GenBank Accession No: NC_000023) 정보에서 염기서열 #60664 (C/T), 염기서열 #60751 (C/T), 그리고 염기서열 #60766 (A/G)에 위치하는 것을 특징으로 하는 성별 판별용 SNP 마커.

청구항 4

모든 인종 집단에서 고도의 다형성을 가지고 있는 개체식별용 30개 SNP 마커로서, 상기 30개의 SNP 마커(GenBank Accession No at dbSNP)는 rs7532151, rs4846468, rs6751657, rs10185531, rs7652776, rs6443222, rs17497475, rs1350191, rs2565007, rs4607417, rs1790006, rs2813838, rs2267708, rs1293288, rs7849782, rs7907658, rs550840, rs543840, rs734075, rs1151849, rs7328030, rs978511, rs7164801, rs955665, rs4791495, rs4647887, rs1785745, rs7230112, rs8113496 및 rs2327088로 이루어진 군에서 선택된 것을 특징으로 하는 개체 식별용 SNP 마커.

청구항 5

청구항 1 내지 청구항 3 중 어느 한 항에 따른 성별 판별용 SNP 마커 6개와 청구항 4항에 따른 개체 식별용 SNP 마커 30개를 조합하는 것을 특징으로 하는 성별과 개체를 동시에 식별할 수 있는 성별 판별 및 개체 식별 방법.

청구항 6

청구항 5에 있어서, 상기 성별 판별은 하나 이상의 성별 판별용 마커를 사용하고, 상기 개체 식별은 개체 식별용 SNP 마커를 사용하되, 식별을 원하는 시료의 개수에 따라 100개 시료는 10개 SNP 마커를 사용하고 1000개 시료는 15개 SNP 마커를 사용하며 10,000개 시료는 20개의 개체식별용 SNP 마커를 사용하는 것을 특징으로 하는 성별 판별 및 개체 식별 방법.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 단일염기변이를 갖는 SNP(single nucleotide polymorphism) 마커를 이용한 성별 결정 및/또는 개체 식별용 마커와 이를 이용한 성별 결정 및/또는 개체 식별 방법에 관한 것이다.

배경 기술

[0002] 성별 식별 및 개체 식별은 범죄수사과학에서 많이 사용되는 중요한 유전자 검사방법으로서 다양한 인종에서 다수의 대립유전자형을 갖고 있는 STR(short tandem repeat) 마커가 주로 사용되고 있다. 그러나 STR 마커는 높은 돌연변이율, 한꺼번에 많은 유전형(multiplexing) 정보를 얻기가 어렵고, 대립유전형을 결정하기 위해서는 큰 사이즈의 증폭 산물이 요구되는 등의 문제로 인하여 그 사용이 제한되고 있다.

[0003] 한편, SNP(Single Nucleotide Polymorphism)는 개인 간에 존재하는 두 개의 다른 유전체에서 단일염기서열의

차이에 기인된 유전변이형을 말하며 전체 인간 유전변이형에서 가장 많은 부분 즉, 전체 인간 유전변이의 약 90% 이상을 차지하는 유전변이 마커이다.

[0004] SNP 마커는 기존의 다른 유전변이형 마커들에 비해 보다 촘촘하고 안정한 유전 지도를 작성할 수 있다는 점과 새로운 기술, 특히 DNA 칩을 포함하는 고효율 유전자형 결정기술(high-throughput genotyping technology)에 보다 적합한 특성을 가지고 있다.

[0005] 따라서 높은 돌연변이율 등의 문제점을 지닌 현재 개체 식별용으로 사용되는 STR 마커를 대신할 새로운 SNP 마커의 개발은 많은 장점을 가지고 다양한 분야에 활용도가 클 것으로 예상된다.

발명의 내용

해결 하고자하는 과제

[0006] 본 발명의 목적은 기존 STR 마커의 문제점을 해결할 수 있는 성별 판별용 SNP 마커 및/또는 개체 식별용 SNP 마커를 제공하는 데에 있다.

[0007] 또한, 본 발명의 다른 목적은 성별 판별용 SNP 마커와 개체 식별용 SNP 마커를 이용한 성별 판별 및 개체 식별 방법을 제공하는 데에 있다.

과제 해결수단

[0008] 상기 목적을 달성하기 위하여, 본 발명은 아멜로제닌(Amelogenin) 유전자(AMELX & AMELY)에서 발굴된 성별 판별용 SNP(single nucleotide polymorphism) 마커 3개와 아연 핑거 단백질 유전자(ZFX & ZFY)에서 발굴된 성별 판별용 SNP(single nucleotide polymorphism) 마커 3개를 포함한 총 6개로 이루어진 성별 판별용 SNP 마커를 제공한다.

[0009] 또한, 본 발명은 모든 인종 집단에서 고도의 다형성을 가지고 있는 개체식별용 30개 SNP 마커로서, 상기 30개의 SNP 마커(GenBank Accession No at dbSNP)는 rs7532151, rs4846468, rs6751657, rs10185531, rs7652776, rs6443222, rs17497475, rs1350191, rs2565007, rs4607417, rs1790006, rs2813838, rs2267708, rs1293288, rs7849782, rs7907658, rs550840, rs543840, rs734075, rs1151849, rs7328030, rs978511, rs7164801, rs955665, rs4791495, rs4647887, rs1785745, rs7230112, rs8113496 및 rs2327088로 이루어진 군에서 선택된 것을 특징으로 하는 개체 식별용 SNP 마커를 제공한다.

[0010] 이하, 본 발명을 보다 상세하게 설명한다.

[0011] 1. 성별 판별 SNP 마커

[0012] 기존의 성별을 결정하는 것으로는 아멜로제닌 유전자(AMELX & AMELY)에 존재하는 Insertion/deletion(Indel) DNA 마커를 이용하여 판별하였으나, 본 발명에서는 아멜로제닌 유전자(AMELX & AMELY)와 아연 핑거 유전자(ZFX & ZFY)에 존재하는 단일염기서열의 차이를 이용하여 사람의 성별을 판별하였다.

[0013] ① 아멜로제닌 유전자의 성별 판별용 SNP 마커 발굴

[0014] 우선 단일염기서열 차이를 이용한 성별 판별을 위해 성염색체 X & Y에 각각 존재하면서 유전자의 염기서열이 매우 유사한 아멜로제닌 유전자 2개(AMELX: X 염색체에 존재 & AMELY: Y 염색체에 존재함)의 염기서열이 매우 유사한 부위(도 3a 참조)를 우선 선별하고, 여자 12명과 남자 12명의 DNA 시료를 이용하여 시퀀싱(sequencing) 하여 성별을 완전히 구분할 수 있는 3개의 단일염기서열(도 4a 참조)을 발굴하였다.

[0015] 성별을 판별할 수 있는 3개의 단일염기서열 위치는 NCBI GenBank에 등록된 AMELX 유전자 염기서열(GenBank Accession No: NC_000023) 정보에서 염기서열 #5139 (C/T), 염기서열 #5196 (C/G), 그리고 염기서열 #5232 (C/T)에 위치하고 있으며 단일염기서열의 차이가 성별을 정확히 판별함을 확인하였다. 이때, 구체적인 염기서열 정보는 도 1과 같다.

[0016] ② 아연 핑거 단백질(ZFX & ZFY) 유전자의 성별 판별용 SNP 마커 발굴

[0017] 아멜로제닌 유전자와 동일하게 성염색체 X & Y에 각각 존재하는 아연 핑거 단백질 유전자 2개(ZFX: X 염색체에 존재 & ZFY: Y 염색체에 존재함)의 염기서열이 매우 유사한 부위(도 3b 참조)를 우선 선별하고, 여자 12명과 남자 12명의 DNA 시료를 이용하여 시퀀싱(sequencing) 하여 성별을 완전히 구분할 수 있는 3개의 단일염기서열(도

4b 참조)을 발굴하였다.

- [0018] 성별을 결정할 수 있는 3개의 단일염기서열 위치는 NCBI GenBank에 등록된 ZFX 유전자 염기서열(GenBank Accession No: NC_000023) 정보에서 염기서열 #60664 (C/T), 염기서열 #60751 (C/T), 그리고 염기서열 #60766 (A/G)에 위치하고 있으며 단일염기서열의 차이가 성별을 결정함을 확인하였다. 이때, 구체적인 염기서열 정보는 도 2와 같다.
- [0019] 즉, 사람의 성별을 판별하기 위한 SNP 마커 발굴을 위해 아멜로제닌 유전자(AMELX 및 AMELY)와 아연 핑거 단백질 유전자(ZFX 및 ZFY) 각각에서 X 염색체와 Y 염색체에서 존재하는 두 유전자 간의 다중 염기서열 정렬로 상동성이 높은 영역을 선별한다. PCR 프라이머를 이용하여 X 및 Y 염색체 상의 유전자 간 하나의 뉴클레오타이드 차이를 포함하는 표적 영역 150-250 bp를 증폭시킬 수 있다(프라이머 서열은 도 3a 및 도 3b에서 밑줄 친 부위의 염기서열임).
- [0020] 일실시예로서, 12명의 남성 및 12명의 여성을 포함한 24명의 개체에 대한 염기서열 분석을 통해 아멜로제닌 유전자에서 3개의 단일염기서열 부위 (AMEL-1, -2, -3)와 아연 핑거 단백질 유전자에서 3개의 단일염기서열 부위 (ZF-1, -2, -3)에서 남성과 여성을 완벽하게 구별할 수 있었다(도 4a 및 도 4b 참조).
- [0021] 또한, 본 발명에 따른 6개의 SNP 마커 중 하나 또는 둘 이상의 SNP 마커 세트를 이용하여 지노타이핑(genotyping)함으로써 성별을 판별할 수 있다. 일실시예로서, 다수의 피검체 DNA 시료(총 960명 DNA 시료)에서 상기 선별된 AMEL-1 (C/T) 부위, AMEL-2 (C/G) 부위 및 ZF-1 (C/T) 부위를 지노타이핑하여 예상 성별과 유전자형이 정확히 일치하였다(도 5 참조).
- [0022] 2. 개체 식별용 SNP 마커
- [0023] 기존의 개체 식별 방법은 STR 마커가 사용되고 있으나, 본 발명에서는 SNP 마커를 이용한 개체 식별 방법을 개발하였으며 이를 위하여 개체 식별용 SNP 마커는 다음과 같은 방법을 사용하여 동정되었다.
- [0024] 우선, 총 50만개의 SNP(500K)를 지노타이핑할 수 있는 Affymetrix 회사의 제품(Genome-Wide Human SNP array 5.0)을 사용하여 한국인 2,000명에서 생산된 유전자형 정보를 사용하여 한국인에서 매우 고도로 다형성(minor allele frequency >0.45 이상)을 보이는 SNP 마커를 다수 선별하였다.
- [0025] 이렇게 선별된 SNP 마커는 다시 다른 인종집단(서양인, 아프리카인 & 동양인)에서 모두 고도의 다형성 (minor allele frequency >0.45이상)을 보이는 SNP 마커 56개를 최종 선별하였다. 그리고 선별된 56개의 SNP 마커에서 최종적으로 30개를 선별(표 1)하여 한국인 960명의 DNA 시료를 이용하여 실제 개체식별 능력을 검정한 결과 매우 우수한 개체식별 능력을 갖고 있었다(표 2).
- [0026] 표 1에 나타난 바와 같이 개체 식별용 SNP 마커 30개의 정보(GenBank SNP ID: dbSNP)는 다음과 같다: rs7532151, rs4846468, rs6751657, rs10185531, rs7652776, rs6443222, rs17497475, rs1350191, rs2565007, rs4607417, rs1790006, rs2813838, rs2267708, rs1293288, rs7849782, rs7907658, rs550840, rs543840, rs734075, rs1151849, rs7328030, rs978511, rs7164801, rs955665, rs4791495, rs4647887, rs1785745, rs7230112, rs8113496 & rs2327088.
- [0027] 또한, 본 발명은 상기 성별 판별용 SNP 마커 6개와 상기 개체 식별용 SNP 마커 30개를 조합하여 사용함으로써 성별과 개체를 동시에 식별할 수 있는 성별 판별 및 개체 식별 방법을 제공한다.
- [0028] 상기 성별 판별은 하나 이상의 성별 판별용 마커를 사용하고, 상기 개체 식별은 개체 식별용 SNP 마커를 사용하되, 10개 SNP 마커를 이용하면 100명까지, 15개 SNP 마커는 1,000명까지의 개체를 식별할 수 있으며, 더 나아가, 20개의 SNP 마커를 사용한 경우에는 약 10,000명(실제분석에서 8,842명의 시료에서 검정됨을 확인함)의 개체 모두를 완벽하게 식별할 수 있다.
- [0029] 따라서, 개체 식별용 시료의 수가 10배 증가할 때 마다 개체 식별용 SNP 마커를 5개씩 증가하여 다양한 개수의 조합으로 개체 식별용 DNA 마커 세트를 사용하여 성별 판별 및 개체 식별을 결정할 수 있다.
- [0030] 또한, SNP 마커의 숫자를 증가하여 최대 40개를 사용하면 한국인 집단에서 STR 마커 16개를 사용하는 표준 개체 식별용 마커 세트와 동일한 개체식별 능력을 가지게 된다.
- [0031] 한편 최종적으로 선별된 30개의 SNP 마커는 마커 간의 거리가 최소 50 Mb (약 50 cM) 이상 떨어져 있어 임의로 어떤 조합을 골라도 모두 동일한 식별능력을 가지게 된다.

[0032] 또한, 본 발명에서는 성별 판별, 개체 식별 및 기타 다양한 외형(예, 피부색, 머리색 및 눈동자색 등)에 관련된 유전자형 정보를 유전자 검사에 추가적으로 사용할 수 있도록 하기 위하여 최소 1개에서 최대 386개 SNP 마커까지 동시에 지노타이핑할 수 있는 Illumina 회사의 BeadExpress 시스템을 사용함으로써 향후에 다양한 표현형에 관련된 SNP 마커를 추가할 수 있는 방법까지 확보하게 되었다.

효 과

[0033] 본 발명에 따른 SNP 마커를 이용하여 성별 판별 및 개체 식별을 하게 되면 STR 마커와 달리 낮은 돌연변이율, 분해된 DNA 시료에 유용한 100bp보다 짧은 앰플리콘 크기의 유용성 및 고효율 지노타이핑 기술 등의 장점이 있어 성별 판별 및 개체 식별에 매우 유용하다.

발명의 실시를 위한 구체적인 내용

[0034] 하기 실시예를 통해 본 발명을 보다 상세하게 설명한다. 그러나, 이러한 실시예는 본 발명의 내용을 구체화하기 위한 설명일 뿐 실시예에 의해 본 발명이 한정되는 것은 아니다.

[0035] 이하 실시예에서 사용된 모든 DNA 시료는 표준 DNA 추출법에 따라 혈액 또는 Epstein Barr virus(EBV)-형질전환 B-세포로부터 분리된 것을 사용하였다. 그리고 이하 실험은 서울아산병원 임상시험심사위원회의 승인을 받아 수행되었다.

[0036] <실시예 1> 성별 결정용 SNP 마커의 선별 및 유효성 평가

[0037] 1. 성별 판별용 SNP 마커의 선별

[0038] 성별 판별용 SNP 마커를 동정하기 위하여, 12개 CEPH 가계 시료(시료고유번호: GM06994, GM07022, GM07347, GM07357, GM12962, GM12960, GM07000, GM07056, GM07346, GM07345, GM12863, GM12861)의 DNA[Coriell Institute for Medical Research(<http://www.coriell.org/>)로부터 구입하여 사용함]와 12개의 한국인 개체의 DNA 시료[질병관리본부 국립보건연구원 유전체센터 유전자원은행팀으로부터 제공받아 사용함]를 사용하였다.

[0039] 아멜로제닌 유전자(AMELX 및 AMELY)와 아연 핑거 단백질 유전자(ZFX 및 ZFY)의 서열 분석을 위하여, GenBank(<http://www.ncbi.nlm.nih.gov/>) 데이터베이스로부터 게놈 서열을 얻고 PCR 프라이머를 Primer3 소프트웨어(<http://frodo.wi.mit.edu/>)를 이용하여 선정하여 X 및 Y 염색체 상의 유전자 간 하나의 염기서열 차이를 포함하는 높은 상동성 표적 영역 150-250 bp를 증폭시켰다(도 3a 및 도 3b 참조). 이때, PCR 프라이머 서열은 밑줄친 부분이다.

[0040] 한편, PCR 증폭 반응은 20 ng 게놈 DNA, 0.15 μ M 각 프라이머, 100 μ M 각 dNTP, 1 unit의 AmpliTaq Gold™(Applied Biosystems, Foster City, CA, USA) 및 1.5mM MgCl₂를 포함한 반응완충액을 함유한 20 μ l의 용액에서 수행되었다. PCR은 94℃에서 5분간 반응 후, 94℃에서 45초, 55℃에서 30초, 72℃에서 1분 총 35 주기를 수행하며, 마지막으로 72℃에서 10분간 연장반응을 수행하였다. PCR 산물(1 μ l)는 시퀀싱용 주형으로 직접 사용하였고, ABI3730 DNA 염기서열분석기 상에서 BigDye® terminator v3.1 cycle sequencing kit(Applied Biosystems, Foster City, CA, USA)를 사용하여 제조자 지침에 따라 시퀀싱하였다. DNA 다형은 PolyPhred 프로그램(<http://droog.gs.washington.edu/PolyPhred.html>)을 사용하여 확인하였다.

[0041] 그 결과, 12명의 남성 및 12명의 여성을 포함한 24명의 개체에 대한 서열 비교를 통해 아멜로제닌 유전자(AMEL-1, -2, -3)와 아연 핑거 단백질 유전자(ZF-1, -2, -3)에서 각각 3개 (총 6개)의 단일염기서열 차이에 의해 남성과 여성을 완벽하게 구별할 수 있었다(도 4a 및 도 4b 참조).

[0042] 2. 성별 판별용 SNP 마커의 유효성 평가

[0043] 앞서 선정된 성별 판별용 SNP 마커의 유효성 검사를 위하여, 총 6개의 성별 결정용 단일염기 부위 중에서 3개의 부위(AMEL-1, AMEL-2 & ZF-1)를 선별하여 조사를 하였다. 본 검정실험은 성별에 대한 정보를 확보한 한국인 DNA 시료(n=960)에서 Illumina VeraCode GoldenGate 분석을 이용하여 각 SNP 마커의 지노타이핑을 제조자의 지침에 따라 수행하였다. 지노타이핑 군집 및 콜링(calling)을 BreadStudio 소프트웨어를 이용하여 수행하였다. 그 결과, AMEL-1, AMEL-2, ZF-1를 포함한 3개 마커의 유전자형 검사는 예상된 성별 유전자형과 완벽하게 일치하였다(도 5 참조).

[0044] <실시예 2> 개체 식별용 SNP 마커의 선별 및 유효성 평가

[0045] 1. 개체 식별용 SNP 마커의 선별

[0046] 개체 식별용 SNP 마커를 선별하기 위하여, 먼저 affymetrix 회사의 500K SNP chip을 이용한 2,000명의 한국인 SNP 지노타이핑 결과 및 4개의 HapMap 인종 시료(일본인, 한족, 아프리카인, 유럽인)의 대립유전자형 빈도로 구성된 Affymetrix 500K SNP 데이터(Nat. Genet. 41, 527-534, 2009)를 이용하였다.

[0047] 다음과 같은 SNP 선별 기준을 근거로 후보 SNP 마커를 선별하였다. 즉, (i) 모든 인종 군에서 마이너 대립유전자 빈도(minor allele frequency, MAF) > 0.45, (ii) 이형접합성(heterozygosity) > 0.45, (iii) 하디-와인버그 평형(Hardy-Weinberg equilibrium, HWE) P-값 > 0.1, (iv) 지노타입 콜 레이트(call rate) = 100%, 그리고 (v) SNP 마커 간 물리적 거리 > 50Mb. 상기 SNP 선별 기준을 근거로 하여 30개의 SNP를 유효성 평가를 위해 선별하였다(표 1).

표 1

SNP (rs#)	Chr.	Position	Allele_1/2	Korean (n=2,000)				Chinese	Japanese	European	African
				Freq_1	Heterozygosity	HWE_P	Match Probability	Freq_1	Freq_1	Freq_1	Freq_1
rs7532151	1	89161532	A/C	0.492	0.500	0.868	0.375	0.489	0.478	0.470	0.510
rs4846468	1	216462550	C/T	0.502	0.500	0.857	0.375	0.523	0.544	0.510	0.540
rs6751657	2	33258655	C/T	0.508	0.500	0.377	0.375	0.465	0.512	0.511	0.530
rs10185531	2	106105267	C/T	0.479	0.499	0.678	0.375	0.533	0.456	0.470	0.500
rs7652776	3	2716024	C/G	0.467	0.498	0.800	0.376	0.478	0.533	0.540	0.490
rs64432222*	3	9138267	A/G	0.481	0.499	0.704	0.375	0.477	0.476	0.479	0.479
rs17497475	4	18851337	A/C	0.455	0.496	0.932	0.377	0.489	0.522	0.530	0.480
rs1350191	4	155026774	A/G	0.486	0.500	0.991	0.375	0.467	0.455	0.479	0.489
rs2565007	5	53856024	A/C	0.458	0.496	0.557	0.377	0.533	0.456	0.490	0.470
rs4607417	6	42086252	C/T	0.531	0.498	0.412	0.376	0.533	0.456	0.510	0.500
rs1790006	6	162297818	A/G	0.470	0.498	0.204	0.376	0.478	0.511	0.500	0.540
rs2813838	7	24115817	C/G	0.457	0.496	0.514	0.377	0.511	0.456	0.500	0.531
rs2267708	7	124179748	C/T	0.511	0.500	0.607	0.375	0.489	0.533	0.550	0.520
rs1293288	8	11755937	A/G	0.482	0.499	0.611	0.375	0.467	0.489	0.530	0.450
rs7849782	9	103467085	C/G	0.466	0.498	0.803	0.376	0.489	0.456	0.470	0.540
rs7907658	10	92782368	A/C	0.524	0.499	0.552	0.376	0.511	0.544	0.510	0.520
rs550840	11	63958215	C/T	0.473	0.499	0.999	0.376	0.500	0.500	0.490	0.520
rs543840	11	115269696	C/T	0.525	0.499	0.271	0.376	0.547	0.500	0.480	0.540
rs734075	12	4367308	A/C	0.469	0.498	0.585	0.376	0.467	0.467	0.521	0.490
rs1151849	12	119824982	A/G	0.454	0.496	0.352	0.377	0.477	0.478	0.480	0.500
rs7328030	13	111067445	A/C	0.517	0.499	0.248	0.375	0.533	0.533	0.540	0.450
rs978511	14	61773670	C/G	0.455	0.496	0.463	0.377	0.456	0.511	0.470	0.520
rs7164801	15	22608408	G/T	0.503	0.500	0.324	0.375	0.500	0.489	0.460	0.480
rs955665	15	77603212	C/T	0.535	0.498	0.303	0.376	0.500	0.533	0.510	0.542
rs4791495	17	12130882	A/G	0.502	0.500	0.532	0.375	0.467	0.533	0.450	0.510
rs4647887	17	72070401	A/G	0.532	0.498	0.245	0.376	0.477	0.544	0.550	0.550
rs1785745	18	21671131	A/G	0.486	0.500	0.993	0.375	0.542	0.465	0.543	0.511
rs7230112	18	71610565	C/T	0.535	0.498	0.467	0.376	0.500	0.534	0.480	0.510
rs8113496	19	34512369	A/G	0.482	0.499	0.919	0.375	0.467	0.489	0.540	0.467
rs2327088	20	8714882	A/G	0.491	0.500	0.940	0.375	0.522	0.522	0.450	0.480
CPI (combined power of identity)							1.76828E-13				
CPD (combined power of discrimination)							0.9999999999982300000 (1 in 5.65522E+12)				

[0048]

[0049] 2. 개체 식별용 SNP 마커의 유효성 평가

[0050] 앞서 선정된 개체 식별용 30개 SNP 마커의 유효성 검사를 위하여, 두 가지 다른 방법을 활용하였다. 첫 번째 검정방법은 실제로 개체에 대한 정보를 확보한 한국인 960명의 DNA 시료(n=960)를 이용하여 선별된 30개의 SNP 마커를 지노타이핑하여 개체식별 능력을 조사하였다. 또 다른 방법은 선별된 30개의 SNP 마커 정보를 가지고 있는 한국인 8,842명의 Affymetrix Genome-wide Human SNP array 5.0 데이터를 분량받아 총 8,842명의 한국인 시료에서 30개의 SNP 마커를 사용하여 개체식별 능력을 검정하였다.

[0051] 우선 실제 한국인 960명의 DNA 시료 세트에서 Illumina회사의 VeraCode GoldenGate 지노타이핑 방법을 사용하여 SNP 마커의 지노타이핑을 제조자의 지침에 따라 수행하였다. 지노타이핑 군집 및 콜링(calling)을 BreadStudio 소프트웨어를 이용하여 수행되었다.

[0052] 선별된 30개의 SNP 마커 중에서, 29개의 SNP 마커가 성공적으로 지노타이핑 되었고, 나머지 하나의 마커(rs6443222)는 지노타이핑이 되지 않았다. 각 SNP 마커의 대립유전자형 및 유전자형 빈도는 표 1과 거의 동일한 것으로 나타났다. 예상한 대로, 선별된 모든 SNP 마커는 고른 대립유전자형 빈도를 나타내었다. 29개 SNP의 총 평균일치확률(combined power of identity)은 4.83×10^{-13} 이며, 이는 총 식별능력(combined power of discrimination) 0.99999999999517과 일치하였다.

[0053] 그리고 960명에서의 지노타이핑 결과에서 15개의 SNP만을 사용한 경우에도 단지 한쌍(2명의 개체)만이 동일한 유전자형을 나타내었는데 이는 15개의 SNP 마커를 사용하면 약 1,000명까지의 개체도 식별할 수 있음을 의미한다.

[0054] 또한, 8,842명 개체를 포함한 Affymetrix SNP array 5.0 유전자형 데이터 분석을 통해서 개체 식별 능력을 분석한 결과에 의하면, 선별된 30개 SNP의 결합 평균일치확률은 1.78×10^{-13} 이며, 이는 결합 식별력 0.99999999999822와 일치하였다. 그리고 표 2에 나타난 바와 같이 개체 식별에 사용되는 SNP 마커의 개수가 5개 증가할 때마다 개체 식별 능력이 10배 증가함을 나타내었다. 즉, 10개 또는 15개 SNP 마커를 사용하면 각각 100명과 1,000명의 개체를 식별할 수 있다. 또한, 20개의 SNP 마커를 사용한 경우에는 8,842명의 개체 모두를 완벽하게 식별할 수 있었다.

표 2

SNP 개수	분석 시료 개수(n)	동일한 유전자형을 나타내는 시료 개수*	평균 일치도(%)
10	100	0-0-2	0.67
	1,000	54-38-39	4.37
	8,842	2,769	31.32
15	100	0-0-0	0.00
	1,000	0-0-2	0.07
	8,842	20	0.23
20	100	0-0-0	0.00
	1,000	0-0-0	0.00
	8,842	0	0.00

[0056] * 100개 및 1,000개 시료의 경우에는 분석 시료를 모든 시료(n=8,842)로부터 무작위로 3회씩 선별하였다.

도면의 간단한 설명

[0057] 도 1은 AMELX 유전자의 유전체 염기서열 NC_000023의 부분염기서열과 성별 판별 염기서열의 위치를 나타낸 것이고,

[0058] 도 2는 ZFX 유전자의 유전체 염기서열 NC_000023의 부분염기서열과 성별 판별 염기서열의 위치를 나타낸 것이고,

[0059] 도 3a 및 도 3b는 각각 아멜로제닌 유전자 및 아연 핑거 단백질 유전자의 게놈 DNA 서열을 정렬한 것이고,

[0060] 도 4a 및 도 4b는 성별 결정용 각 아멜로제닌 유전자 및 아연 핑거 단백질 유전자에 있어서 SNP 동정 결과를 나타낸 것이고,

[0061] 도 5는 960명 개체에서 GoldenGate Assay을 이용한 AMEL-2 마커의 지노타이핑 결과를 나타낸 것이다(자주색 점은 남성, 파란색 점은 여성을 나타냄).

도면

도면1

5101 gccaatggta aacctgecte tetgtttctc accagtiac[C/T]c ttectatggt tacgagccca
 5161 tgggtggatg gctgcaccac caaatcatcc ccgtg[C/G]tgte ccaacagcac cccccgactc
 5221 acacctgca g[C/T]ctcateac cacatcccag tgggtgccage tcagcagccc gtgatecccc
 5281 agcaaccaat gatgcccggt cctggccaac actccatgac tccaatccaa caccaccage

도면2

60541 ccgtgtact gactgtgatt aactaccaa caagaagata agtttacaca accacctgga
 60601 gagccacaag ctgaccagca aggccagaaa ggccattgaa tgcgatgagt gtgggaagca
 60661 ttt[C/T]tctcat gcaggggctt tgtttactca caaatgggtg cataaggaaa aaggagccaa
 60721 caaatgcac aagtgtaat tctgtgaata [C/T]gagacagct gaaca[A/G]gggt
 tattgaatcg
 60781 ccacctcttg gcagtcacaa gcaagaactt tctcatatt tgtgtggagt gtgtaaggg

도면3a

AMELX	ACTGCTGCTTCTCTGGTTGGAGTCACCTGAGCCAATGGTAAACCTGCCCTCTCTGTTTCTC	180
AMELY	ACTGCTGCTTCTCTGGTTGGAGTCACCTGAGCCAATGGTAAACCTGCATCTCTGTTTCTC	60

AMELX	ACCAGTACCTTTCCTATGGTTACGAGCCCATGGGTGGATGGCTGCACCAACCAATCATEC	240
AMELY	ACCAGTACTTTCCTATGGTTACGAGCCCATGGGTGGATGGCTGCACCAACCAATCATEC	120

AMEL-1		
AMELX	CCGTGCTGTCCCAACAGCACCCCTGACTCACACCCCTGCAGCTCATCACCACATCCAG	300
AMELY	CCGTGCTGTCCCAACAGCACCCCTGACTCACACCCCTGCAGCTCATCACCACATCCAG	180

AMEL-2		
AMELX	TGGTGCCAGCTCAGCAGCCC	
AMELY	TGGTGCCAGCTCAGCAGCCC	

AMEL-3		

도면3b

```

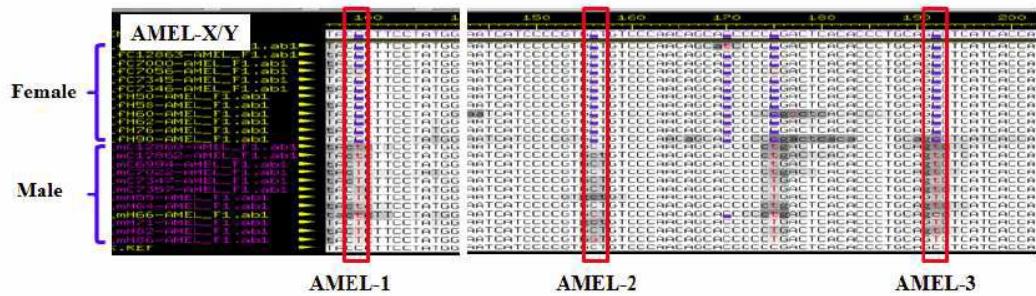
ZFX      ACCAACAAAGAAGATAAGTTTACACAAC CACCTGGAGAGCCACAAGCTGACCAGCAAGGCA 58720
ZFY      ACCAATAAGAAGATAAGTTTACATAAC CACCTGGAGAGCCACAAGCTGACCAGCAAGGCA 43523
*****

ZFX      GAGAAGGCCATTGAATGCGATGAGTGTGGGAAGCATTTCCTCTCATGCAGGGGCTTTGTTT 58780
ZFY      GAGAAGGCCATTGAATGTGATGAGTGTGGGAAGCATTTCCTCTCATGCAGGGGCTTTGTTT 43583
*****
                                ZF-1
ZFX      ACTCACAAAATGGTGCATAAGGAAAAAGGAGCCAACA AATGCACAAGTGTA AATTCGT 58840
ZFY      ACTCACAAAATGGTGCATAAGGAAAAAGGGGCCAACA AATGCACAAGTGTA AATTCGT 43643
*****

ZFX      GAATACGAGACAGCTGAACAAGGGTTATTGAATCGCCACCTCTTGGCAGTCCACAGCAAG 58900
ZFY      GAATATGAGACAGCTGAACAAGGGTTATTGAATCGCCACCTCTTGGCAGTCCACAGCAAG 43703
*****
                ZF-2                ZF-3
ZFX      AACCTTCCTCATATTTGTGTGGAGTGTGGTAA
ZFY      AACCTTCCTCATATTTGTGTGGAGTGTGGTAA
*****

```

도면4a



Genotype	AMEL-1 (C/T) AMEL-2 (C/G) AMEL-3 (C/T)			
	Female (XX; N=12)	12-0-0	12-0-0	12-0-0
	Male (XY; N=12)	0-12-0	0-12-0	0-12-0
Seq. Images	Female (XX)			
	Male (XY)			

도면4b



Genotype	ZF-1 (C/T)	ZF-2 (C/T)	ZF-3 (A/G)
	Female (XX; N=12) 12-0-0	12-0-0	12-0-0
Seq. Images	Female (XX)	Female (XX)	Female (XX)
	Male (XY)	Male (XY)	Male (XY)

도면5

