



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0077658
(43) 공개일자 2013년07월09일

(51) 국제특허분류(Int. Cl.)
G06F 19/10 (2011.01)
(21) 출원번호 10-2011-0146492
(22) 출원일자 2011년12월29일
심사청구일자 2011년12월29일

(71) 출원인
대한민국(관리부서: 행정안전부 국립과학수사연구원장)
서울특별시 양천구 지양로 139 (신월동)
(72) 발명자
한면수
경기도 수원시 장안구 조원동 한일타운 145동
1904호
이양한
서울특별시 종로구 신영동 105-4 201호
(뒷면에 계속)
(74) 대리인
특허법인 웰-엘엔케이

전체 청구항 수 : 총 11 항

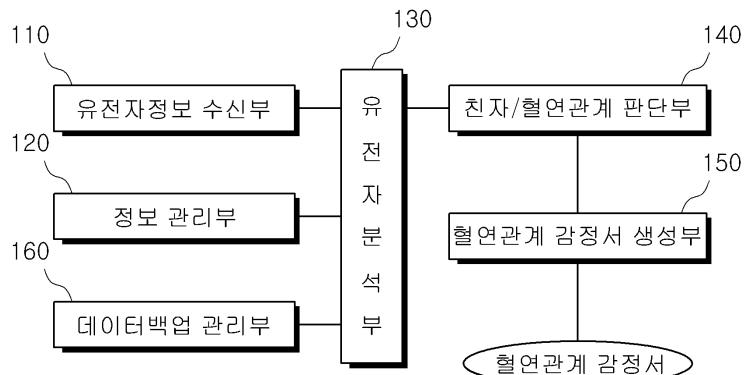
(54) 발명의 명칭 유전자 분석을 이용한 친자확인 검증시스템 및 그 방법

(57) 요 약

친자확인을 위한 부계지수 계산 알고리즘을 구현하여 친자확인 검증의 신뢰성과 정확성을 확보할 수 있도록 한 유전자 분석을 이용한 친자확인 검증시스템 및 그 방법이 개시된다.

개시된 유전자 분석을 이용한 친자확인 검증시스템은, 유전자 검사장비로부터 대립유전자 검사 결과를 업로드 받는 유전자 정보 수신부와; 내부 메모리에 저장된 친자확인 검증 알고리즘을 인출하고, 상기 인출한 친자확인 검증 알고리즘에 따라 상기 수신한 대립유전자 검사 결과를 대응시켜 친자확인 및 혈연 관계 확인을 위한 친부 확률 및 혈연관계 확률을 추출하는 유전자 분석부와; 상기 유전자 분석부에서 추출한 친부 확률 및 혈연관계 확률을 기반으로 친자 및 혈연 관계를 판단하는 친자/혈연관계 판단부와; 상기 친자/혈연관계 판단부에서 판단한 친자 및 혈연관계 판단 결과를 기반으로 혈연관계 감정서를 생성하는 혈연관계 감정서 생성부를 구비한다.

대 표 도 - 도1



(72) 발명자

김남예

서울특별시 양천구 신정6동 목동신시가지아파트
806동 1106호

김옥

충청남도 천안시 서북구 불당동 호반리젠시빌아파트
106동 1204호

조재형

경기도 성남시 분당구 수내동 파크타운서안아파트
112동 902호

김기철

충청남도 천안시 동남구 안서동 226-24 203호

구교찬

충청남도 천안시 동남구 안서동 e-편한세상대림아
파트 108동 1504호

특허청구의 범위

청구항 1

유전자 검사장비로부터 대립유전자 검사 결과를 업로드 받는 유전자 정보 수신부와;

내부 메모리에 저장된 친자확인 검증 알고리즘을 인출하고, 상기 인출한 친자확인 검증 알고리즘에 따라 상기 수신한 대립유전자 검사 결과를 대응시켜 친자확인 및 혈연 관계 확인을 위한 친부 확률 및 혈연관계 확률을 추출하는 유전자 분석부와;

상기 유전자 분석부에서 추출한 친부 확률 및 혈연관계 확률을 기반으로 친자 및 혈연 관계를 판단하는 친자/혈연관계 판단부와;

상기 친자/혈연관계 판단부에서 판단한 친자 및 혈연관계 판단 결과를 기반으로 혈연관계 감정서를 생성하는 혈연관계 감정서 생성부를 포함하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증시스템.

청구항 2

청구항 1에 있어서, 상기 친자확인 검증 알고리즘은,

유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하며, 상기 추출한 대립유전자의 빈도를 친부지수를 산출하기 위한 친부지수 공식에 적용하여 각각의 유전자형에 대한 친부지수를 산출하고, 상기 각각의 유전자형에 대한 친부지수를 기반으로 연산하여 결합친부지수(CPI)를 산출하며, 상기 산출한 결합친부지수로 친부확률을 산출하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증시스템.

청구항 3

청구항 2에 있어서, 상기 친자확인 검증 알고리즘은,

혈연관계 확인을 위한 Identity by descent(IBM) 상수를 산출하고, 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하며, 상기 산출한 IBM 상수와 대립유전자 빈도를 혈연관계 지수계산 공식에 적용하여 각각의 마커 값을 산출하고, 상기 산출한 마커 값을 기반으로 혈연관계 지수를 산출하고, 상기 산출한 혈연관계 지수를 혈연관계 확률 계산식에 적용하여 혈연관계 확률을 산출하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증시스템.

청구항 4

청구항 1에 있어서,

분석할 유전자의 관계 정보, 유전자 분석 결과 정보, 생성된 혈연관계 감정서를 저장하여 관리하는 데이터 백업 관리부를 더 포함하며,

상기 유전자 분석부는 대립유전자의 빈도를 추출하기 위한 대립유전자 빈도표, 표준 STR 마커가 저장된 데이터 베이스를 포함하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증시스템.

청구항 5

(a) 유전자 검사장비로부터 유전자 검사 결과를 업로드 받는 단계와;

(b) 상기 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하는 단계와;

- (c) 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하는 단계와;
- (d) 상기 추출한 대립유전자의 빈도를 친부지수를 산출하기 위한 친부지수 공식에 적용하여 각각의 유전자형에 대한 친부지수를 산출하는 단계와;
- (e) 상기 각각의 유전자형에 대한 친부지수를 기반으로 연산하여 결합친부지수(CPI)를 산출하는 단계와;
- (f) 상기 산출한 결합친부지수로 친부확률을 산출하는 단계를 포함하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증방법.

청구항 6

청구항 5에 있어서, 상기 마커는 FBI 표준 마커인 13 CODIS 마커를 포함하며, 국제적으로 표준화되어 사용화된 Identifiler, PowerPlex 16 등에 포함된 17개 마커를 표준 마커로 이용하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증방법.

청구항 7

청구항 5에 있어서, 상기 결합친부지수는 각각의 유전자형에 대한 친부지수를 승산하여 산출하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증방법.

청구항 8

청구항 5에 있어서, 상기 친부확률(PP(%))은 하기와 같은 수식에 의해 산출하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증방법.

$$PP(\%) = \frac{CPI \times 0.1}{CPI \times 0.1 + (1 - 0.1)} \times 100$$

여기서 CPI는 결합친부지수를 나타낸다.

청구항 9

청구항 5에 있어서,

- (g) 혈연관계 확인을 위한 Identity by descent(IBM) 상수를 산출하는 단계와;
- (h) 상기 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하는 단계와;
- (i) 상기 산출한 IBM 상수와 대립유전자 빈도를 혈연관계 지수계산 공식에 적용하여 각각의 마커 값을 산출하고, 상기 산출한 마커 값을 기반으로 혈연관계 지수를 산출하는 단계와;
- (j) 상기 산출한 혈연관계 지수를 혈연관계확률 계산식에 적용하여 혈연관계확률을 산출하는 단계를 더 포함하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증방법.

청구항 10

청구항 9에 있어서, 상기 혈연관계지수는,

상기 산출한 마커 값을 모두 승산하여 승산치를 산출하고, Unrelated의 곱해진 값으로 상기 승산치를 제산하여 산출하는 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증방법.

청구항 11

청구항 9에 있어서, 상기 혈연관계확률 계산식은

$$\text{혈연관계확률} = \left\{ \frac{\text{혈연관계지수}}{(\text{혈연관계지수}+1)} \right\} \times 100$$

인 것을 특징으로 하는 유전자 분석을 이용한 친자확인 검증방법.

명세서

기술분야

[0001] 본 발명은 유전자(DNA) 분석을 이용한 친자확인 검증시스템 및 그 방법에 관한 것으로서, 더욱 상세하게는 친자 확인을 위한 부계지수 계산 알고리즘을 구현하여 친자확인 검증의 신뢰성과 정확성을 확보할 수 있도록 한 유전자 분석을 이용한 친자확인 검증시스템 및 그 방법에 관한 것이다.

배경기술

[0002] 최근 실종자, 변사자, 미아 등의 유전자 분석 자료는 지속적으로 증가하고 있다. 또한, 과학수사뿐만 아니라 사회 및 법률적인 문제와 의료분야에 이르기까지 광범위한 분야에서 개인식별 및 친자확인에 관한 과학적 검증을 필요로 하고 있다.

[0003] 현재 국제적으로 검증된 DNA 분석 자료, 특히 STR 마커(Short Tandem Repeat Marker)를 이용한 친자확인 분석 시스템이 다양한 분야에서 이용되고 있다.

[0004] 여기서 유전자란 생물체의 개개의 유전 형질을 발현시키는 원인이 되는 인자를 말한다. 유전자는 일정한 순서로 배열되며, 생식 세포를 통하여 부모로부터 자손에게 유전 정보를 전달하는 역할을 한다. 유전자의 본체는 DNA이며, 단백질의 합성과 합성에 필요한 단백질의 종류를 지령함으로써 체내에서 일어나는 대부분의 화학반응을 매개하는 역할을 한다. 즉, DNA에서 세포나 조직의 구성성분이 순서대로 생성될 수 있도록 시기와 장소를 프로그램하고, 그 생물의 생활사를 통해 생물의 활동성을 결정하며, 그 생물 특유의 개성을 결정하는데 필요한 정보인 유전정보를 저장하고 있는 부분이 유전자(gene)인 것이다.

[0005] 영국의 유전학자인 Alec Jeffreys는 인체의 DNA 중에서 형질유전은 하지 않고, 16개의 염기를 단위서열로 적렬 반복되는 유전자를 발견하였는데, 이 반복부위를 Variable Number of Tandem Repeats(VNTR)라 한다. 이 부분의 염기 서열 반복 횟수는 사람마다 판이하여, 이를 이용해 개인 식별이 가능하다는 사실을 밝혀냈다. Jeffreys는 VNTR의 다형성이 사람마다 지문이 다른 것과 유사하다는 점에 착안해 '유전자지문(DNA Fingerprint)'이라는 용어를 처음으로 사용하였다. 이와 같은 유전자의 특성에 기인하여 개인의 식별, 특정한 질병 또는 소인(素因)의 검사 등의 목적으로 혈액·모발·타액 등의 검사대상물로부터 염색체·유전자 등을 분석하는 행위를 '유전자검사'라 하며, 특히 개인 식별 등의 목적으로 유전자 중 유전정보를 포함하지 않는 부분에 대하여 분석을 한 결과물이 바로 '유전자정보'이다.

[0006] Jeffreys가 인간 유전자의 다형성을 밝혀내고, 이를 이용해 개인을 식별 해내는 방법을 고안한 이후에 범죄 수사는 한 단계 발전을 하게 된다. 그는 이 방법을 이용해 실제로 강간살인사건을 해결하여 그 효용성을 증명하기도 했다.

[0007] 그러나 Jeffreys의 유전자 분석방법은 그 기술상의 문제로 인하여 보편화되지는 못하였다. 하지만, 이후에 다양한 분석 기법들이 개발되면서 인간 유전자의 다형성을 이용하여 개인을 식별하려는 시도가 결실을 거두게 되었다.

[0008] 개인 식별의 목적으로 유전자의 다형성을 밝히기 위해 처음 사용된 방법은 1980년 Wyman과 White에 의해 소개된 제한효소절편길이다형성(Restriction Fragment Length Polymorphism, RFLP)방법이다. 이 방법은 제한효소(Restriction Enzyme, RE)와 전기영동, 그리고 Southern blotting을 이용하는 방법인데, 우선 인체 유전자를

EcoRI라 부르는 제한효소로 잘라낸 후 이를 '탐식자'로 하여 여기에 다시 방사선 동위원소처리를 하고 제한효소 처리된 피검사자의 유전자에 붙여보는 방법이다. 사람의 유전자 중에는 탐식자에 결합하는 특정 부위가 있으며 이 부위는 사람마다 차이가 있다는 사실을 밝혀냄에 따라 이를 이용해 동일인 여부를 식별하려는 방법이다. 탐식자의 종류에 따라 단좌 탐식자에 의한 RFLP법과 다좌 탐식자에 의한 RFLP법으로 구분된다.

- [0009] 이 방법은 다형성이 높다는 면에서는 현재 시행되고 있는 여러 검사 방법들도 가지지 못하는 장점이 있음에도 변형되지 않는 다양한 시료가 필요하고, 시험방법이 까다롭고 적지 않은 비용과 시간이 소모되며, 분석 결과를 표준화하기 어렵다는 점 등의 문제점이 지적되어 현재는 범죄 수사 목적의 증거물을 검사에는 사용되지 않고 있다.
- [0010] 1985년 미국의 Saiki 등은 Taq polymerase라는 효소를 이용하여 인간유전자의 특정부위를 인체 밖에서 증폭할 수 있는 방법을 개발하고 이를 중합효소 연쇄반응(Polymerase Chain Reaction, PCR)이라고 명명하였다. 유전자에 중합효소 복제 지점을 지정하는 시발체(primer), 유전자 합성재료인 뉴클레오타이드(nucleotide, dNTP)를 넣어 적절하게 온도를 변화시키면 유전자의 특정 부위가 매우 높은 비율로 증가하게 된다. 유전자의 특정 부위를 체외에서 증폭할 수 있게 됨에 따라 여러 분야의 유전자 연구에서 혁신적인 발전이 이루어졌다. 특히 법의학적으로 미량의 시료를 다루는 경우가 많은데 시료의 양적 제한점을 극복할 수 있게 되어 이전에는 검사 결과를 얻지 못하였던 시료의 경우에도 검사할 수 있게 되었고, 시료가 부패한 경우에도 분석할 수 있게 되었다. PCR증폭 산물을 간단하게 전기영동 한 후 크기를 비교하면 유전자형이 결정되며, 이와 같은 기법을 증폭산물 길이다형성(Amp-FLP)법이라 한다.
- [0011] PCR법에 사용되는 유전자는 STR유전자인데, 상염색체 상에 존재하고 2 ~ 4개의 염기가 반복되는 특성이 있으며, 이 반복되는 수가 개인마다 다르기 때문에 개인 식별에 이용된다. 또한, 상염색체 상의 STR 좌위는 부모로부터 각각 유전되기 때문에 개인 식별뿐만 아니라 신원확인에도 사용될 수 있다. 한편, 성염색체에도 STR좌위가 존재하는데 이 경우 Y-STR은 부계의 가족관계를 분석하는 데에도 사용 가능하다.
- [0012] PCR 분석방법은 자료의 오염으로 1/100만 이하의 다른 DNA가 함유된 경우에 원치 않는 염기서열의 증폭으로 인해 오류가 발생하는 문제점이 지적되기도 하지만, 절차의 간편성, 짧은 분석시간, 높은 분석 민감성이라는 장점 때문에 독일, 일본, 우리나라 등에서 형사사법 분야에 이용되고 있다.
- [0013] 다음으로, 미토콘드리아 DNA(mt-DNA) 분석법은 기존의 방법과는 달리 핵 외부에 존재하는 미토콘드리아에 존재하는 DNA를 분석하는 것이다. 이 미토콘드리아 DNA는 핵 DNA와는 다르게 모계로부터만 유전이 되기 때문에, 형제자매만 있는 경우의 신원확인 등에 사용되며, 한 세포 안에 많은 수의 복제수를 가지고 있고, 원형의 작은 유전자로 되어있기 때문에 치아나 유골과 같이 오래되었거나 변질된 시료, 모근이 없는 머리카락이나 손톱 등과 같이 STR분석이 곤란하거나 불가능한 시료의 분석에 주로 사용된다.
- [0014] 유전자정보를 분석하기 위해 DNA 샘플을 채취해야하는 데, 여기서 DNA를 채취하는 방법에는 여러 가지가 있다. 의료 목적이나 친자확인 등의 목적으로 DNA 샘플을 수집하는 방법으로 채혈이나 조직 적출 등의 방법이 주로 사용된다. 범죄 수사 목적의 유전자감식 절차는 일반적으로 우선 증거물을 확보하고, 이 증거물로부터 DNA를 분리하여, 분석을 위한 처리를 거친 뒤에 결과를 판독하는 절차가 유전자감식의 핵심이라 할 수 있다.
- [0015] 유전자의 다형성을 이용하여 개인을 식별해 내려면 우선 검체에서 깨끗한 유전자 상을 얻어야 하고 이와 비교할 대조 유전자 상을 확보해야 한다. 깨끗한 유전자상은 검체에서 추출되는 유전자의 질과 양에 좌우되므로 최대한의 유전자를 추출할 수 있도록 증거물 채취에 섬세한 주위를 기울여야 하며 채취된 증거물은 실제검사가 시행될 때까지 더 이상의 유전자 변질이 일어나지 않도록 보관에 주의해야한다. 사건 현장이나 피해자에게 잔류하고 있는 혈흔과 혈액에서 DNA 증거를 채취하기 위해서는 혈액이 건조되지 않은 상태로 보관하면 혈액이 부패 될 수 있으므로 주의를 기울여야한다. 따라서 통째로 수거되는 의류나 종이와 같은 증거물은 음지에서 건조한 후 보관해야 하며, 통째로 수거하기 어려운 경우에는 면봉을 이용하여 혈액을 떼어내 건조한 후 수거해야 한다.
- [0016] 이후 채취한 DNA 샘플을 유전자 분석하게 되는 데, 유전자 분석은 여러 단계를 거쳐 이루어지며, 통상 1 ~ 3주가 걸리는 것으로 알려져 있다.
- [0017] 우리나라에서도 유전자정보는 광범위한 분야에서 활용되고 있다. 주로 개인 식별 목적과 의료목적으로 이용되고 있는데, 특히 유전자형의 다형성을 이용한 개인식별 분야의 중요성이 부각되고 있으며, 그 중에서도 범죄 수사 분야에서의 활용도가 갈수록 증가하고 있다. 유전자 정보가 활용되는 대표적인 분야는 범죄 수사목적, 의료 목적, 미아찾기 등 다양하다.
- [0018] 유전자를 검사하는 방법은 다양하며, 대표적인 것이 친자확인 검사(STR), 부계확인 검사(Y-STR), 모계확인 검사

(X-STR), 모계확인 검사(mtDNA), 개인식별 검사, DNA 프로필 검사, 정액확인, 혈액확인, 유골 검사 등이 있다.

- [0019] 먼저, 친자확인 검사(STR)는 부, 모, 자의 STR 유전자형을 분석하여 부, 모, 자 유전자형의 일치 여부를 확인 후 통계적 분석을 통해 친자확인 관계인지를 판별하는 검사이다. 사람의 몸을 구성하는 세포는 부, 모로부터 각각 23개의 염색체를 물려받아 총 46개의 염색체를 가지게 된다. 친자확인검사는 이러한 유전적 특성을 이용하여 부, 모의 유전자 타입과 자식의 유전자 타입이 서로 일치하는지를 조사하여 일치 가능성을 확률 값으로 나타내 친자 여부를 판별하는 검사이다. 검사결과는 성염색체를 포함한 총 15 ~ 18개의 STR유전자 좌위를 조사하여 부, 모의 유전자 타입과 자의 유전자 타입이 검사된 유전자 좌위에서 모두 일치할 경우 친부(모)확률로 조사하여 친부(모) 확률 값이 99.99% 이상이면 친부(모)로 인정한다. 단, 1 ~ 2개의 유전자 좌위에서 불일치가 나타났을 경우 돌연변이의 가능성을 조사한다.
- [0020] 다음으로, 부계확인 검사(Y-STR)는 조부 → 부 → 자(子)로 전달되는 Y 염색체상의 STR을 검사하여 부계혈연관계를 확인하는 검사이다. 남자의 성(sex)염색체로 XY를 가지는데, 이중 Y염색체는 남자에게만 존재하여 아들로만 전달된다. 아들은 다시 그 아들에게 Y염색체를 유전시키므로 동일 집안의 모든 남자는 동일한 Y염색체를 가지게 된다. 따라서 Y염색체를 이용한 부계확인 검사는 Y염색체에 존재하는 특정 STR 유전자형을 검사하여 부계 혈연관계를 확인하는 검사이다. 검사한 모든 Y-STR 유전자 좌위에서 일치할 경우, 부계혈연관계가 성립된다. 단, 성립된 Y-STR 유전자좌위 중 1 ~ 2개가 불일치할 경우 돌연변이의 가능성을 조사한다.
- [0021] 다음으로, 모계확인 검사(X-STR)는 X염색체상의 STR을 검사하여 여성의 부계혈연관계를 확인하는 검사이다. 남성의 성염색체는 XY이고 여성의 성염색체는 XX이므로, X염색체에 존재하는 STR 유전자 좌위에는 남자는 단일 대립유전자를 가지며 여자는 두 개의 대립유전자를 가지게 된다. 여성의 경우 X염색체상의 두 대립유전자 중 하나는 친모로부터, 나머지 하나는 친부로부터 물려받게 되므로 X염색체상의 존재하는 STR 좌위에서 공유하는 대립 형질을 가지게 된다. 이러한 유전적 성질을 이용하여 X염색체상의 X-STR 분석은 여성의 부계혈연관계를 확인가능하게 한다. 검사한 모든 X-STR 유전자 좌위에서 일치할 경우, 여성의 부계혈연관계가 성립된다. 단, 검사된 X-STR 유전자 좌위 중 1 ~ 2개가 불일치할 경우 돌연변이의 가능성을 조사한다.
- [0022] 다음으로, 모계확인 검사(mtDNA)는 어머니로부터 자식들로 전달되는 특성이 있는 미토콘드리아 DNA를 검사하여 모계혈연관계를 확인하는 검사이다. 사람의 세포 내에는 크게 2종류의 DNA가 존재한다. 하나는 세포핵 내에 존재하는 DNA로 세포의 대부분을 차지하고 있으며, 나머지는 아주 작은 일부분으로 세포질 내의 미토콘드리아에 존재하는 DNA이다. 세포 내에서 수정이 이루어질 때 정자는 난자를 뚫고 세포핵의 DNA만을 전달하고 미토콘드리아는 DNA를 전달하지 않는다. 따라서 미토콘드리아 DNA의 유전자정보는 전적으로 난자, 즉 어머니의 미토콘드리아 DNA에 의존하게 되며 독특한 모계유전현상이 일어나게 된다. 이러한 모계로만 유전되는 특징이 있는 미토콘드리아 DNA는 외할머니에게서 어머니와 어머니의 남녀 형제들에게로, 그리고 어머니에서 또 아들, 딸로 전해진다. 따라서 모계 유전되는 미토콘드리아 DNA를 분석하면 같은 모계 혈족인지를 확인할 수 있다. 미토콘드리아 DNA의 괴변이 지역인 HV1, HV2, HV3 지역의 염기서열분석을 통하여 검사된 모든 염기서열이 일치할 경우 동일모계가 성립함을 인정하다. 단, 검사된 모든 염기 중 1 ~ 2개가 불일치할 경우 돌연변이의 가능성을 조사한다.
- [0023] 다음으로, 개인식별 검사는 사건현장의 검체가 서로 일치하는지를 조사하여 동일인의 것인지를 판별하는 검사이다. DNA는 개인의 모든 유전정보를 담고 있는 유전물질로서 모든 사람들은 부모로부터 DNA를 물려받아 다른 사람들과 구별되는 독특한 유전자 패턴인 DNA 프로필을 갖게 된다. DNA 프로필이란 사람을 식별할 수 있는 그 사람의 공유한 유전정보로서 연쇄반복이 나타나는 유전자 좌위들의 조합이다. 개인식별 검사는 두 개 이상의 검체가 유전자적 증거인 DNA프로필이 서로 일치하는지를 조사하여 동일인의 것인지를 판별하는 검사이다. 예를 들면, 사건 현장에 떨어진 가해자의 머리카락과 용의자의 검체에 대해 DNA검사를 수행한 후 검사된 유전자 좌위에서 서로 일치하는지를 조사하여 동일인 것 인지를 판별하는 검사이다. 두 개의 검체에서 DNA 분석을 통해 검사된 모든 유전자 좌위에서 유전자타입이 모두 일치할 경우 동일인의 것으로 판명한다. 단 1개 이상의 유전자 좌위에서 서로 불일치가 나타날 경우, 서로 다른 사람의 것임을 나타낸다.
- [0024] 다음으로, DNA 프로필 검사는 개인(1인)의 유전자 감식을 통해 불의의 사고 시 신원 확인에 이용할 수 있으며, 유아의 경우 미아 발생 시 찾을 수 있는 강력한 식별력을 지닌 검사이다. 개인(1인)의 DNA 프로필 검사를 통해 불의의 사고 시 신원확인에 이용할 수 있으며 유아의 경우, DNA 프로필검사를 통해 미리 DNA 프로필을 알고 있으면 미아 발생 시 유전정보의 데이터베이스를 통해 미아 찾기가 가능하다.
- [0025] 여기서 주지한 바와 같은 모든 유전자 검사 방법들은, 부계지수 등 친자 및 혈연관계 확인을 위한 통계학적 계산을 대부분 수기에 의하거나 엑셀을 통한 수작업으로 수행하게 된다.

발명의 내용

해결하려는 과제

- [0026] 그러나 상기와 같은 종래의 다양한 유전자 검사 방법들은 부계지수 등 친자 및 혈연관계 확인을 위한 통계학적 계산이 대부분 수기에 의하거나 엑셀을 통한 수작업으로 이루어지기 때문에 매우 불편하고, 검사 시간이 많이 소요되는 단점이 있었다.
- [0027] 또한, 종래에 개발된 친자확인 및 혈연관계 확인을 위한 소프트웨어는, 해외에서 개발된 것으로 비용이 비싸고 현재 우리나라에서 보편적으로 사용하고 있는 시스템과 호환시키는 것도 어렵다는 단점이 있다.
- [0028] 이에 본 발명은 상기와 같은 종래 기술에서 발생하는 제반 문제점을 해결하기 위해서 제안된 것으로서,
- [0029] 본 발명이 해결하고자 하는 과제는 친자확인을 위한 부계지수 계산 알고리즘을 구현하여 친자확인 검증의 신뢰성과 정확성을 확보할 수 있도록 한 유전자 분석을 이용한 친자확인 검증시스템 및 그 방법을 제공하는 데 있다.
- [0030] 본 발명이 해결하고자 하는 다른 과제는 대량의 유전자 분석 자료를 체계적으로 관리하고 효과적으로 비교 검색 할 수 있는 유전자 분석을 이용한 친자확인 검증시스템 및 그 방법을 제공하는 데 있다.
- [0031] 본 발명이 해결하고자 하는 또 다른 과제는 유전자 검사장비에서 검사한 결과(유전자 프로필)를 자동화된 시스템을 통해서 입력받도록 함으로써 입력과정에서의 실수에 의한 오류를 방지하고, 사용이 간편하며 국제적으로 표준화되고 신뢰성 있는 통계학적 분석을 통한 친자 검증이 가능한 유전자 분석을 이용한 친자확인 검증시스템 및 그 방법을 제공하는 데 있다.

과제의 해결 수단

- [0032] 상기와 같은 과제들을 해결하기 위한 본 발명의 바람직한 실시 예에 따른 "유전자 분석을 이용한 친자확인 검증 방법"은,
- [0033] (a) 유전자 검사장비로부터 유전자 검사 결과를 업로드 받는 단계와;
- [0034] (b) 상기 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하는 단계와;
- [0035] (c) 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하는 단계와;
- [0036] (d) 상기 추출한 대립유전자의 빈도를 친부지수를 산출하기 위한 친부지수 공식에 적용하여 각각의 유전자형에 대한 친부지수를 산출하는 단계와;
- [0037] (e) 상기 각각의 유전자형에 대한 친부지수를 기반으로 연산하여 결합친부지수(CPI)를 산출하는 단계와;
- [0038] (f) 상기 산출한 결합친부지수로 친부확률을 산출하는 단계를 포함하는 것을 특징으로 한다.
- [0039] 상기에서 마커는 FBI 표준 마커인 13 CODIS 마커를 포함하며, 국제적으로 표준화되어 상용화된 Identifier, PowerPlex 16 등에 포함된 17개 마커를 표준 마커로 이용하는 것을 특징으로 한다.
- [0040] 상기에서 결합친부지수는 각각의 유전자형에 대한 친부지수를 승산하여 산출하는 것을 특징으로 한다.
- [0041] 상기에서 친부확률(PP(%))은 하기와 같은 수식에 의해 산출하는 것을 특징으로 한다.

$$PP(\%) = \frac{CPI \times 0.1}{CPI \times 0.1 + (1 - 0.1)} \times 100$$

- [0043] 여기서 CPI는 결합친부지수를 나타낸다.
- [0044] 또한, 본 발명에 따른 "유전자 분석을 이용한 친자확인 검증방법"은,
- [0045] (g) 혈연관계 확인을 위한 Identity by descent(IBD) 상수를 산출하는 단계와;
- [0046] (h) 상기 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하는 단계와;
- [0047] (i) 상기 산출한 IBD 상수와 대립유전자 빈도를 혈연관계 지수계산 공식에 적용하여 각각의 마커 값을 산출하고, 상기 산출한 마커 값을 기반으로 혈연관계 지수를 산출하는 단계와;
- [0048] (j) 상기 산출한 혈연관계 지수를 혈연관계확률 계산식에 적용하여 혈연관계확률을 산출하는 단계를 더 포함하는 것을 특징으로 한다.
- [0049] 상기에서 혈연관계지수는,
- [0050] 상기 산출한 마커 값을 모두 승산하여 승산치를 산출하고, Unrelated의 곱해진 값을으로 상기 승산치를 제산하여 산출하는 것을 특징으로 한다.
- [0051] 상기에서 혈연관계 확률 계산식은

$$\text{혈연관계 확률} = \left\{ \frac{\text{혈연관계지수}}{(\text{혈연관계지수}+1)} \right\} \times 100$$
 인 것을 특징으로 한다.
- [0052] 상기와 같은 과정들을 해결하기 위한 본 발명의 바람직한 실시 예에 따른 "유전자 분석을 이용한 친자확인 검증 시스템"은,
- [0053] 유전자 검사장비로부터 대립유전자 검사 결과를 업로드 받는 유전자 정보 수신부와;
- [0054] 내부 메모리에 저장된 친자확인 검증 알고리즘을 인출하고, 상기 인출한 친자확인 검증 알고리즘에 따라 상기 수신한 대립유전자 검사 결과를 대응시켜 친자확인 및 혈연 관계 확인을 위한 친부 확률 및 혈연관계 확률을 추출하는 유전자 분석부와;
- [0055] 상기 유전자 분석부에서 추출한 친부 확률 및 혈연관계 확률을 기반으로 친자 및 혈연 관계를 판단하는 친자/혈연관계 판단부와;
- [0056] 상기 친자/혈연관계 판단부에서 판단한 친자 및 혈연관계 판단 결과를 기반으로 혈연관계 감정서를 생성하는 혈연관계 감정서 생성부를 포함하는 것을 특징으로 한다.
- [0057] 상기 친자확인 검증 알고리즘은,
- [0058] 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하며, 상기 추출한 대립유전자의 빈도를 친부지수를 산출하기 위한 친부지수 공식에 적용하여 각각의 유전자형에 대한 친부지수를 산출하고, 상기 각각의 유전자형에 대한 친부지수를 기반으로 연산하여 결합친부지수(CPI)를 산출하며, 상기 산출한 결합친부지수로 친부확률을 산출하는 것을 포함하는 것을 특징으로 한다.
- [0059] 바람직하게 상기 친자확인 검증 알고리즘은,
- [0060] 혈연관계 확인을 위한 Identity by descent(IBD) 상수를 산출하고, 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출

하며, 상기 산출한 IBD 상수와 대립유전자 빈도를 혈연관계 지수계산 공식에 적용하여 각각의 마커 값들을 산출하고, 상기 산출한 마커 값들을 기반으로 혈연관계 지수를 산출하고, 상기 산출한 혈연관계 지수를 혈연관계화를 계산식에 적용하여 혈연관계화률을 산출하는 것을 더 포함하는 것을 특징으로 한다.

발명의 효과

- [0061] 본 발명에 따르면 구현된 친자확인을 위한 부계지수 계산 알고리즘으로 친자확인 검증의 신뢰성과 정확성을 확보할 수 있는 장점이 있다.
- [0062] 또한, 본 발명에 따르면 대량의 유전자 분석 자료를 체계적으로 관리하고 효과적으로 비교 검색할 수 있는 장점이 있다.
- [0063] 또한, 유전자 검사장비에서 검사한 결과(유전자 프로필)를 자동화된 시스템을 통해서 입력받도록 함으로써 입력 과정에서의 실수에 의한 오류를 방지하고, 사용에 편의성을 제공해주는 장점도 있다.
- [0064] 또한, 본 발명에 따르면 실종자, 변사자 및 미아찾기 등에 활용 가능한 가족 검색 및 검증 시스템을 제공해주는 장점이 있다.
- [0065] 또한, 본 발명에 따르면 많은 양의 유전자정보를 쉽게 관리하고, 친자/가족관계를 빠르게 분석하여 실종자 및 변사자의 가족을 빠르게 찾아줄 수 있는 장점도 있다.

도면의 간단한 설명

- [0066] 도 1은 본 발명에 따른 유전자 분석을 이용한 친자확인 검증시스템의 구성도.
- 도 2는 본 발명에 따른 유전자 분석을 이용한 친자확인 검증방법을 보인 흐름도.
- 도 3은 본 발명에 적용된 표준 STR 마커 예시도.
- 도 4는 본 발명에서 부계 및 모계지수 계산을 위한 공식 예시도.
- 도 5는 본 발명에서 친부지수 계산을 위한 샘플 예시도.
- 도 6은 본 발명에서 마커별 대립유전자의 빈도 예시도.
- 도 7은 본 발명에서 STR마커를 이용한 친자관계 분석 결과도.
- 도 8은 본 발명에서 적용되는 IBD 상수 예시도.
- 도 9는 본 발명에서 혈연관계 확인을 위한 공식 예시도.
- 도 10은 본 발명에서 혈연관계 지수계산을 위한 샘플 예시도.
- 도 11은 본 발명에서 STR 마커를 이용한 혈연관계 분석 결과도.
- 도 12는 본 발명에서 데이터베이스 ERD 예시도.
- 도 13은 본 발명에서 친자관계 분석결과 판정 룰(RULE) 예시도.

발명을 실시하기 위한 구체적인 내용

- [0067] 이하 본 발명의 바람직한 실시 예를 첨부한 도면에 의거 상세히 설명하면 다음과 같다. 본 발명을 설명하기에 앞서 관련된 공지 기능 혹은 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그에 대한 상세한 설명은 생략한다.
- [0068] 도 1은 본 발명의 바람직한 실시 예에 따른 "유전자 분석을 이용한 친자확인 검증시스템"의 구성도로서, 유전자 정보 수신부(110), 정보 관리부(120), 유전자 분석부(130), 친자/혈연관계 판단부(140), 혈연관계 감정서 생성부(150) 및 데이터 백업 관리부(160)로 구성된다.
- [0069] 여기서 유전자 정보 수신부(110)는 유전자 검사장비(예를 들어, GeneMapper)로부터 대립유전자 검사 결과를 업로드 받는 역할을 하는 것으로서, 도면에는 도시하지 않았지만 통신을 위한 통신 모듈을 포함하는 것이 바람직

하다. 특히, 웹 기반으로 친자 확인 검증을 수행하기 위해서 인터넷 등과 연결되어 데이터를 수신받는 것이 바람직하다.

- [0070] 정보 관리부(120)는 분석할 유전자의 관계 정보를 관리하는 역할을 하는 것으로서, 사고자 정보라든지 가족 정보 등을 입력받고, 이를 유전자 정보 수신부(110)에서 업로드 한 대립유전자 검사 결과와 매칭하여 관리하는 역할을 하며, 이러한 정보 관리부(120)를 통해 실험 키트를 입력할 수 있도록 하여 추후 키트의 추가 및 관리를 할 수 있도록 한다.
- [0071] 유전자 분석부(130)는 내부 메모리에 저장된 친자확인 검증 알고리즘을 인출하고, 상기 인출한 친자확인 검증 알고리즘에 따라 상기 유전자 정보 수신부(110)를 통해 수신한 대립유전자 검사 결과를 대응시켜 친자확인 및 혈연 관계 확인을 위한 친부 확률 및 혈연관계 확률을 추출하는 역할을 한다.
- [0072] 여기서 유전자 분석부(130)는 데이터베이스를 포함하고, 이러한 데이터베이스에는 대립유전자의 빈도를 추출하기 위한 대립유전자 빈도표, 표준 STR 마커 등이 저장된다. 도 12는 본 발명에 적용되는 데이터베이스의 ERD를 나타낸 것이다.
- [0073] 상기 친자확인 검증 알고리즘은, 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하며, 상기 추출한 대립유전자의 빈도를 친부지수를 산출하기 위한 친부지수 공식에 적용하여 각각의 유전자형에 대한 친부지수를 산출하고, 상기 각각의 유전자형에 대한 친부지수를 기반으로 연산하여 결합친부지수(CPI)를 산출하며, 상기 산출한 결합친부지수로 친부확률을 산출하는 과정을 포함한다.
- [0074] 또한, 바람직하게 상기 친자확인 검증 알고리즘은, 혈연관계 확인을 위한 Identity by descent(IBM) 상수를 산출하고, 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하며, 상기 산출한 IBM 상수와 대립유전자 빈도를 혈연관계 지수계산 공식에 적용하여 각각의 마커 값을 산출하고, 상기 산출한 마커 값을 기반으로 혈연관계 지수를 산출하고, 상기 산출한 혈연관계 지수를 혈연관계확률 계산식에 적용하여 혈연관계확률을 산출하는 과정을 더 포함한다.
- [0075] 친자/혈연관계 판단부(140)는 상기 유전자 분석부(130)에서 추출한 친부 확률 및 혈연관계 확률을 기반으로 친자 및 혈연 관계를 판단하는 역할을 하게 되며, 혈연관계 감정서 생성부(150)는 상기 친자/혈연관계 판단부(140)에서 판단한 친자 및 혈연관계 판단 결과를 기반으로 혈연관계 감정서를 생성하는 역할을 한다.
- [0076] 여기서 혈연관계 감정서는 기본적으로 정해진 감정서 서식(포맷)에 친부 확률 정보, 혈연관계 확률 정보 등을 입력할 수 있도록 한 서식이다.
- [0077] 데이터 백업 관리부(160)는 상기 정보 관리부(120)에서 생성된 분석할 유전자의 관계 정보, 유전자 분석 결과 정보, 생성된 혈연관계 감정서 등을 저장하여 관리하기 위한 역할을 한다.
- [0078] 도 2는 본 발명의 바람직한 실시 예에 따른 "유전자 분석을 이용한 친자확인 검증방법"을 보인 흐름도로서, S는 단계(Step)를 나타낸다.
- [0079] 이에 도시된 바와 같이, (a) 유전자 검사장비로부터 유전자 검사 결과를 업로드 받는 단계(S101)와; (b) 상기 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하는 단계(S102)와; (c) 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하는 단계(S103)와; (d) 상기 추출한 대립유전자의 빈도를 친부지수를 산출하기 위한 친부지수 공식에 적용하여 각각의 유전자형에 대한 친부지수를 산출하는 단계(S104)와; (e) 상기 각각의 유전자형에 대한 친부지수를 기반으로 연산하여 결합친부지수(CPI)를 산출하는 단계(S105)와; (f) 상기 산출한 결합친부지수로 친부확률을 산출하는 단계(S106)와; (g) 혈연관계 확인을 위한 Identity by descent(IBM) 상수를 산출하는 단계(S107)와; (h) 상기 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하는 단계(S108)와; (i) 상기 산출한 IBM 상수와 대립유전자 빈도를 혈연관계 지수계산 공식에 적용하여 각각의 마커 값을 산출하고, 상기 산출한 마커 값을 기반으로 혈연관계 지수를 산출하는 단계(S109)와; (j) 상기 산출한 혈연관계 지수를 혈연관계확률 계산식에 적용하여 혈연관계확률을 산출하는 단계(S110)를 포함한다.
- [0080] 이와 같이 구성되는 본 발명에 따른 "유전자 분석을 이용한 친자확인 검증시스템 및 그 방법"은, 친자확인을 위한 부계지수 계산 알고리즘 및 실종자와 변사자, 미아의 가족을 찾기 위한 검색 알고리즘을 제공하고, 제공된 알고리즘을 활용하여 친자확인 및 실종자, 변사자와 미아를 찾고 사용자 중심의 웹 기반 알고리즘을 통해 통계

학적인 검증을 수행하게 된다.

- [0081] 이를 위해 본 발명에서는 친자, 혈연관계 분석을 위한 표준 STR 마커를 선정하여, 유전자 분석부(130) 내의 테이터베이스에 저장하고, 계산시 이러한 마커를 활용하게 된다.
- [0082] 도 3은 본 발명에 적용되는 표준 STR 마커로서, FBI 표준 마커인 13 CODIS 마커를 포함하여, 국제적으로 표준화되어 상용화되고 있는 Identifiler, PowerPlex 16 등에 포함된 17개 마커를 표준 STR 마커로 선정하였다. Identifiler를 기준으로 하고, PowerPlex 16의 Penta D와 Penta E 마커를 부가적으로 포함하였다. 또한, 향후 식별력을 높이기 위해 새로운 키트 시스템의 추가 및 관리를 가능하게 함으로써 새로운 마커를 계산에 이용할 수 있도록 하였다.
- [0083] 이와 같이 표준 STR 마커를 선정하여 데이터베이스에 저장한 상태에서, 유전자 정보 수신부(110)를 통해 분석하고자 하는 유전자 정보(유전자 검사 결과)를 유전자 검사장비로부터 업로드 받게 된다(S101).
- [0084] 이렇게 업로드 받은 유전자 정보는 정보 관리부(120)에서 검사 대상 관련자 정보와 매칭되어 유전자 분석부(130)에 입력되고, 유전자 분석부(130)는 내부 메모리에 저장된 친자확인 검증 알고리즘을 추출한 후, 상기 친자확인 검증 알고리즘을 통해 상기 입력되는 유전자 검사결과를 분석하게 된다.
- [0085] 분석 과정으로서 첫 번째로는 상기 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하게 된다(S102).
- [0086] 여기서 본 발명에서는 친자확인 및 친자확인 결과의 통계학적인 검증을 위해, 도 4에 도시한 바와 같이 한쪽 부모와 자녀(duo), 양쪽 부모와 자녀(trio) 사이에 부계/모계지수를 계산하게 되는 데, 이때 계산을 위한 공식은 Forensic DNA Evidence Interpretation 10장(2005, Buckleton 등, CRC Press)과 Lee 등(Forensic Sci Int. 2000. 114: 57-65.)을 이용하였다. 돌연변이의 경우 지수계산에는 Brenner's method를 적용하였다.
- [0087] 여기서 STR마커 중 마커가 D3S1358일 경우, 도 5에 도시한 바와 같이, 두 대립유전자(Allele)는 자(15, 16)이고 부(15, 16)이므로, a=15, b=16이다.
- [0088] 다음으로, 상기 결정한 유전자형을 도 6과 같은 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하게 된다(S103). 여기서 대립유전자의 빈도표도 미리 데이터베이스에 저장해 놓고 활용하는 것이 바람직하다.
- [0089] 상기 대립유전자의 빈도표를 참조하여 각 대립유전자의 빈도를 찾아보면, $P_{15}=0.4$, $P_{16}=0.321$ 임을 알 수 있다.
- [0090] 다음으로, 상기 추출한 대립유전자의 빈도를 친부지수를 산출하기 위한 친부지수(도 4의 PI) 공식에 적용하여 각각의 유전자형에 대한 친부지수를 산출하게 된다(S104).
- [0091] 도 4를 참조하면 Genotype Child가 ab이고, Genotype Alleged Father가 ab이므로, 친부지수(Paternity Index, PI)의 공식인 $(P_a+P_b)/(4P_aP_b)$ 를 적용하면 $(0.4+0.321)/(4 \times 0.4 \times 0.321) = 1.403$ 임을 구할 수 있다.
- [0092] 이후 상기 각각의 유전자형에 대한 친부지수를 기반으로 연산하여 결합친부지수(CPI)를 산출하게 된다(S105). 예컨대, 산출한 각 마커의 PI 값을 승산하여 결합친부지수(Combined Paternity Index, CPI)를 산출하게 된다.
- [0093] 이어 상기 산출한 결합친부지수(CPI)로 친부확률(Probability of Paternity, PP)(친자확률이라고도 함)을 산출하게 된다(S106). 여기서 친부확률을 산출하는 공식은 하기 수식 1과 같다.

수학식 1

$$PP(\%) = \frac{CPI \times 0.1}{CPI \times 0.1 + (1 - 0.1)} \times 100$$

- [0094]
- [0095] 도 7은 PowerPlex16에 이용된 STR마커의 유전자정보를 이용해 구해진 친부지수 및 친자확률이며, 친자확률(PP)이 99.99%이다.
- [0096] 이렇게 하여 친자확률이 산출되면, 그 산출된 친자확률 값을 친자/혈연관계 판단부(140)에 전달되고, 상기 친자/혈연관계 판단부(140)는 상기 전달된 친자확률 값을 미리 설정된 기준 판정 룰(RULE)에 적용하여 친자 여부를 판정하게 된다. 도 13은 친자관계 분석결과 판정 룰의 예시이다. 여기서 도 7에 도시된 친자 확률은 99.99%이므로

로 검사 대립유전자는 친부자지간으로 판정할 수 있다.

- [0097] 상기와 같이 친자 여부를 판단한 후에는 해당 결과가 혈연관계 감정서 생성부(150)에 전달되고, 상기 혈연관계 감정서 생성부(150)는 도 7과 같이 미리 생성된 서식에 각각의 정보를 기록하여 혈연관계 감정서를 생성하게 된다.
- [0098] 이렇게 생성되는 혈연관계 감정서는 도면에는 도시하지 않았지만 데이터 백업 관리부(160)를 통해 데이터베이스에 저장되어 관리됨과 동시에 프린터 등으로 전달되어, 혈연관계 감정서가 문서로 출력된다.
- [0099] 한편, 본 발명에서는 상기와 같은 과정을 통해 친자 여부를 판단할 수 있으나, 혈연관계확인 및 혈연관계 확인 결과의 통계학적인 검증을 수행하여 더욱 정확하게 혈연관계를 분석하게 된다.
- [0100] 즉, 혈연관계 확인을 위한 Identity by descent(IBD) 상수를 산출하게 된다(S107). 여기서 IBD 상수는 도 8과 같은 상수 산출 시트를 통해 수행하게 되는 데, 이러한 IBD 상수 산출 시트도 미리 실험을 통해 제작하여 유전자 분석부(130) 내의 데이터베이스에 저장해 놓고 활용하는 것이 바람직하다.
- [0101] 다음으로, 상기 유전자 검사 결과에 포함된 마커 값을 기반으로 유전자형을 결정하고, 상기 결정한 유전자형을 도 6과 같은 대립유전자의 빈도표에 대입하여 대립유전자의 빈도를 추출하게 된다(S108).
- [0102] 이어 상기 산출한 IBD 상수와 대립유전자 빈도를 도 9에 도시한 혈연관계 지수계산 공식(Frequency)에 적용하여 각각의 마커 값을 산출하고, 상기 산출한 마커 값을 기반으로 혈연관계 지수를 산출하게 된다(S109).
- [0103] 여기서 혈연관계 지수계산 방법은 친자관계 지수계산 방법과 동일하게 두 대립유전자를 비교하여 빈도를 구하고, 도 9를 이용하여 혈연관계지수를 구할 수 있다. 단, 혈연관계와 친자관계 지수계산에서 차이점은 Parent-Child, Full sibs, Half sibs, First cousins, Unrelated에 따라 구분하여 각 공식과 Φ 를 적용하여 도 10과 같이 마커별로 각 값을 산출한다.
- [0104] 여기서 혈연관계지수는 상기 산출한 마커 값을 모두 승산하여 승산치를 산출하고, Unrelated의 곱해진 값으로 상기 승산치를 제산하여 산출하게 된다.
- [0105] 이후 상기 산출한 혈연관계 지수를 아래에 기재한 수식 2와 같은 혈연관계확률 계산식에 적용하여 혈연관계확률을 산출하게 된다(S110).

수학식 2

$$\text{혈연관계확률} = \left\{ \frac{\text{혈연관계지수}}{(\text{혈연관계지수} + 1)} \right\} \times 100$$

[0106]

- [0107] 도 11은 PowerPlex16에 이용된 STR마커의 유전자정보를 이용해 구해진 혈연관계지수 및 혈연관계확률이며, 도 11을 참조하면 Parent-Child의 확률이 가장 높고, 99.99%이므로 친부자지간으로 판정할 수 있다.
- [0108] 이렇게 하여 혈연관계확률이 산출되면, 그 산출된 혈연관계확률 값을 친자/혈연관계 판단부(140)에 전달되고, 상기 친자/혈연관계 판단부(140)는 상기 전달된 혈연관계확률 값을 미리 설정된 기준 판정 룰(RULE)에 적용하여 친자 여부를 판정하게 된다. 도 13은 친자관계 분석결과 판정 룰의 예시이다. 여기서 도 11에 도시된 혈연관계 확률 중 Parent-Child의 확률이 가장 높고, 99.99%이므로 친자관계로 판정할 수 있다.
- [0109] 상기와 같이 친자 여부를 판단한 후에는 해당 결과가 혈연관계 감정서 생성부(150)에 전달되고, 상기 혈연관계 감정서 생성부(150)는 도 7과 같이 미리 생성된 서식에 각각의 정보를 기록하여 혈연관계 감정서를 생성하게 된다.
- [0110] 이렇게 생성되는 혈연관계 감정서는 도면에는 도시하지 않았지만 데이터 백업 관리부(160)를 통해 데이터베이스에 저장되어 관리됨과 동시에 프린터 등으로 전달되어, 혈연관계 감정서가 문서로 출력된다.
- [0111] 이러한 본 발명에 따르면 친자확인 및 혈연관계 지수 계산 알고리즘을 활용하여, 폭발적으로 증가하는 대량의 실종자, 변사자, 미아에 대한 상 염색체 STR 유전자 분석 자료를 이용하여 친자확인 및 혈연관계를 효율적으로 수행할 수 있게 되는 것이다. 아울러 유전자 정보는 GeneMapper로부터 출력된 파일을 자동으로 입력하여 사용함으로써 대량의 정보도 손쉽게 입력할 수 있는 장점이 있으며, 입력된 대상 간에 일대일 비교를 통해 친자/혈족

관계 여부를 정확하게 확인할 수 있게 된다. 또한, 부분 시료(임의 마커)를 선택하여 비교하거나 데이터베이스에 입력된 전체 시료(전체 마커)를 비교하여 관계검증을 수행할 수 있으며, 향후 새로운 키트와 이에 해당되는 마커를 추가할 수 있으므로 시스템의 확장성도 향상시켰다.

- [0112] 본 발명은 상술한 특정의 바람직한 실시 예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변형실시가 가능한 것은 물론이고, 그와 같은 변경은 청구범위 기재의 범위 내에 있게 된다.

부호의 설명

[0113] 110… 유전자 정보 수신부

120… 정보 관리부

130… 유전자 분석부

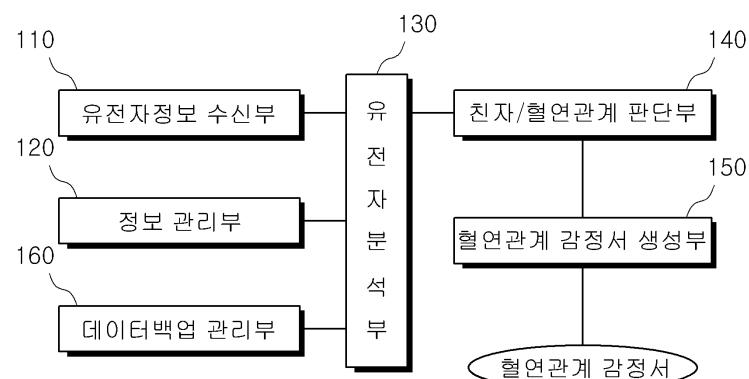
140… 친자/혈연관계 판단부

150… 혈연관계 감정서 생성부

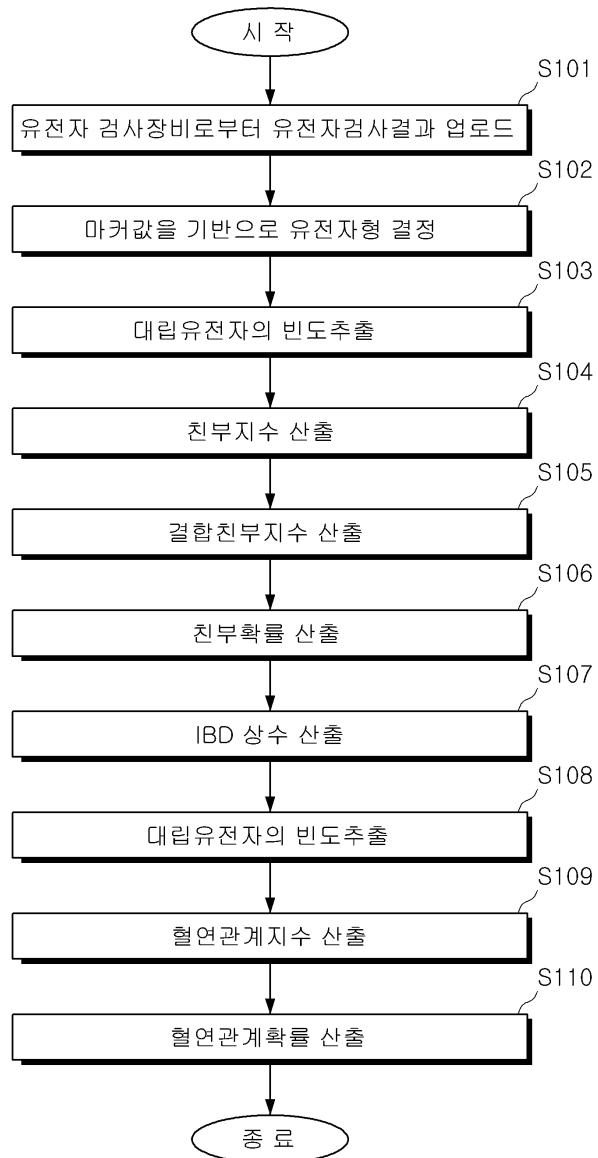
160… 데이터 백업 관리부

도면

도면1



도면2



도면3

Identifier와 PowerPlex 16에서 이용된 STR 마커와 FBI CODIS 마커

CODIS	Identifier	PowerPlex16
VWA	VWA	VWA
TPOX	TPOX	TPOX
TH01	TH01	TH01
FGA	FGA	FGA
D3S1358	D3S1358	D3S1358
D5S818	D5S818	D5S818
D7S820	D7S820	D7S820
D8S1179	D8S1179	D8S1179
D13S317	D13S317	D13S317
D16S539	D16S539	D16S539
D18S51	D18S51	D18S51
D21S11	D21S11	D21S11
CSF1PO	CSF1PO	CSF1PO
Amelogenin	D2S1338	Penta D
	D19S433	Penta E
	Amelogenin	Amelogenin

도면4

부계 및 모계지수 계산을 위한 공식

PI for nonexcluded combinations of maternal and paternal genotypes						
Genotype	Mother	Genotype	Child	Genotype	Alleged Father	PI
aa		aa				
ab				aa		$\frac{1}{P_a}$
bb						
bc		ab				
aa						
ab		aa				
ac						
bb						
bc		ab				$\frac{1}{2P_a}$
bc						
cc						
cd		ac				
				aa		$\frac{1}{P_a + P_b}$
				ab		
ab		ab				
				ac		$\frac{1}{2(P_a + P_b)}$

PI for nonexcluded combinations of paternal and child genotypes						
Genotype	Mother	Genotype	Child	Genotype	Alleged Father	PI
				aa		$\frac{1}{P_a}$
				ab		
				ab		$\frac{1}{2P_a}$
				aa		
				ab		$\frac{P_a + P_b}{4P_a P_b}$
				ac		$\frac{1}{4P_a}$

도면5

Makrer	자	부	Genotype Child	Genotype Alleged Father	PI
D3S1358	15, 16	15, 16	ab	ab	$\frac{P_a + P_b}{4P_a P_b}$
THO1	9, 9.3	8, 9	ab	ab	$\frac{1}{4P_a}$

도면6

Allele frequencies and staristicak parameters regarding the 15 STR loci of Korean population(n=165)

Allele	D81358	THO1	D21S11	D18S50	Penta E	D58018	D13S17	D7S820	D165509	CSF1PO	Penta D	vWA	D851179	TTOX	FGA
5	-	-	-	-	0.067	-	-	-	-	-	-	-	-	-	-
6	-	0.112	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	0.298	-	-	-	0.009	0.003	0.003	-	-	-	0.003	-	-	-
8	-	0.048	-	-	0.006	0.003	0.294	0.118	-	-	0.03	-	-	0.449	-
9	-	0.464	-	0.003	0.009	0.073	0.124	0.048	0.294	0.039	0.342	-	0.003	0.121	-
9.3	-	0.064	-	-	-	-	-	-	-	-	-	-	-	-	-
10	0.003	0.015	-	0.009	0.049	0.21	0.136	0.215	0.167	0.239	0.173	-	0.142	0.046	-
11	-	0.006	-	0.009	0.079	0.312	0.221	0.373	0.255	0.242	0.164	-	0.136	0.333	-
12	-	-	-	0.049	0.139	0.242	0.194	0.188	0.194	0.394	0.146	0.003	0.124	0.046	-
13	-	-	-	0.212	0.045	0.136	0.021	0.052	0.085	0.079	0.1	0.021	0.224	0.006	-
13.2	-	-	-	0.003	-	-	-	-	-	-	-	-	-	-	-
13.3	-	0.003	-	-	-	-	-	-	-	-	-	-	-	-	-
14	0.039	-	-	0.188	0.085	0.015	0.006	0.003	0.006	0.006	0.027	0.212	0.152	-	-
15	0.4	-	-	0.173	0.164	-	-	-	-	-	0.018	0.046	0.158	-	-
16	0.321	-	-	0.085	0.076	-	-	-	-	-	-	0.17	0.049	-	-
17	0.182	-	-	0.076	0.064	-	-	-	-	-	-	0.264	0.006	-	-
18	0.055	-	-	0.07	0.073	-	-	-	-	-	-	0.194	0.006	-	0.015
19	-	-	-	0.033	0.045	-	-	-	-	-	-	0.07	-	-	0.052
20	-	-	0.003	0.033	0.042	-	-	-	-	-	-	0.015	-	-	0.055
21	-	-	-	0.024	0.027	-	-	-	-	-	-	-	-	-	0.155
22	-	-	-	0.021	0.018	-	-	-	-	-	-	0.003	-	-	0.197
22.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.003
23	-	-	-	0.009	0.009	-	-	-	-	-	-	-	-	-	0.255
23.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.006
24	-	-	-	0.003	0.003	-	-	-	-	-	-	-	-	-	0.146
24.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.003
25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.073
26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.036
27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.006
28	-	0.042	-	-	-	-	-	-	-	-	-	-	-	-	-
28.2	-	0.006	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	0.252	-	-	-	-	-	-	-	-	-	-	-	-	-
29.2	-	0.006	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	0.318	-	-	-	-	-	-	-	-	-	-	-	-	-
30.2	-	0.009	-	-	-	-	-	-	-	-	-	-	-	-	-
31	-	0.133	-	-	-	-	-	-	-	-	-	-	-	-	-
31.2	-	0.046	-	-	-	-	-	-	-	-	-	-	-	-	-
32	-	0.021	-	-	-	-	-	-	-	-	-	-	-	-	-
32.2	-	0.088	-	-	-	-	-	-	-	-	-	-	-	-	-
33	-	0.015	-	-	-	-	-	-	-	-	-	-	-	-	-
33.2	-	0.058	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	0.003	-	-	-	-	-	-	-	-	-	-	-	-	-

도면7

Marker	사고자	가족	a	b	Pa	Pb	Type	Formula	PI
	10-M-0009-6	10-M-1009-1							
D3S1358	15 / 16	15 / 16	15	16	0.4	0.321	abab	Pa+Pb/4PaPb	1.403816166347042
TH01	9 / 9.3	8 / 9	9	9.3	0.464	0.064	abac	1/4Pa	0.5387931189518625
D21S11	30.2 / 31.2	30.2 / 30.2	30.2	31.2	0.009	0.045	abaa	1/2Pa	55.55555794710005
D18S51	17 / 19	14 / 19	19	17	0.033	0.076	abac	1/4Pa	7.575757616807607
D5S818	11 / 12	11 / 11	11	12	0.312	0.242	abaa	1/2Pa	1.6025640694993966
Penta_E	5 / 13	5 / 7	5	13	0.067	0.045	abac	1/4Pa	3.7313431906365553
D13S317	12 / 13	12 / 12	12	13	0.194	0.021	abaa	1/2Pa	2.577319511610616
D7S820	10 / 11	10 / 10	10	11	0.215	0.373	abaa	1/2Pa	2.3255813566654555
D16S539	9 / 11	9 / 12	9	11	0.294	0.255	abac	1/4Pa	0.890340136744003
CSF1PO	10 / 12	10 / 13	10	12	0.239	0.394	abac	1/4Pa	1.0460251369504043
Penta_D	12 / 15	12 / 12	12	15	0.145	0.018	abaa	1/2Pa	3.4482759612919205
vWA	17 / 18	17 / 18	17	18	0.264	0.194	abab	Pa+Pb/4PaPb	2.2356294579062586
D8S1179	13 / 14	13 / 14	13	14	0.224	0.152	abab	Pa+Pb/4PaPb	2.7608082891140016
TP0X	8 / 11	8 / 8	8	11	0.448	0.333	abaa	1/2Pa	1.1160713941219977
FGA	20 / 25	20 / 25	20	25	0.055	0.073	abab	Pa+Pb/4PaPb	7.970112154663454

CPI (Combined Paternity Index) 1921301.6595981177 PP (Probability of Paternity) 99.99953156976697%

도면8

Identity by descent 상수 $\{\Phi_2, \Phi_1, \Phi_0\}$

Relationship category	Φ_2	Φ_1	Φ_0
Parent-Child	0	1	0
Full sibs	0.25	0.50	0.25
Half sibs, avuncular	0	0.50	0.50
First cousins	0	0.25	0.75
Unrelated	0	0	1
Formula	$[P_2(xy) \times \Phi_2] + [P_1(xy) \times \Phi_1] + [P_0(xy) \times \Phi_0]$		

도면9

혈족관계확인 공식

allele 1	allele 2	Frequency
AB	AB	$\Phi_2 + 0.5\Phi_1(P_A + P_B) + 2\Phi_0P_A P_B$
AA	AA	$\Phi_2 + \Phi_1 P_A + \Phi_0 P_A^2$
AA	AB	$\Phi_1 P_B + 2\Phi_0 P_A P_B$
AB	AA	$0.5\Phi_1 P_A + \Phi_0 P_A^2$
AB	AC	$0.5\Phi_1 P_C + 2\Phi_0 P_A P_C$
AB	CD	$2\Phi_0 P_C P_D$
AA	BB	$\Phi_0 P_B^2$
AA	BC	$2\Phi_0 P_B P_C$
BC	AA	$\Phi_0 P_A^2$

도면10

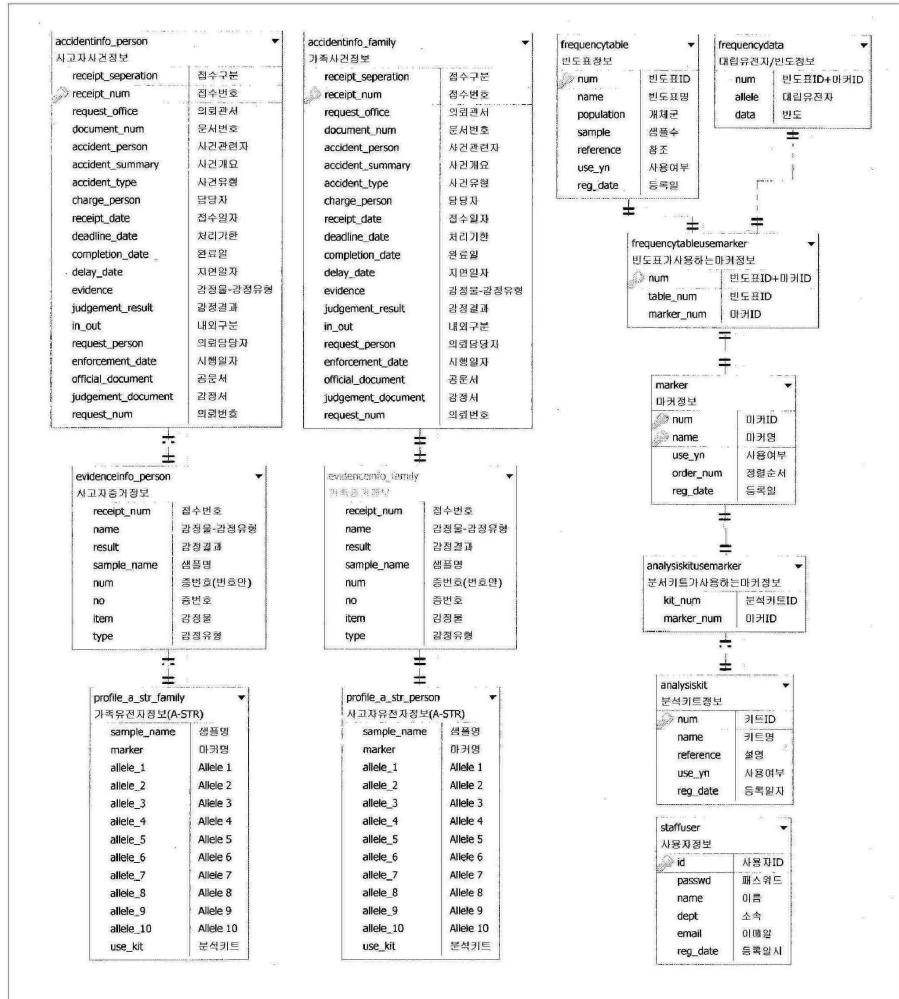
혈연관계 지수계산을 위한 샘플

Person 1	Person 2	Allele frequency
15	15	$P_{15} = 0.4$
16	16	$P_{16} = 0.321$
AB / AB		$\Phi_2 + 0.5\Phi_1(P_A + P_B) + 2\Phi_0 P_A P_B$
Parent-Child		$0.5 \times 1 \times (P_A + P_B)$
Full sibs		$0.25 + 0.5 \times 0.5(P_A + P_B) + 2 \times 0.25 P_A P_B$
Half sibs, avuncular		$0.5 \times 0.5(P_A + P_B) + 2 \times 0.5 P_A P_B$
First cousins		$0.5 \times 0.25(P_A + P_B) + 2 \times 0.75 P_A P_B$
Unrelated		$2 \times 1 \times (P_A \times P_B)$

도면11

Marker	사고자		가족		a (Pa)	b (Pb)	c (Pc)	d (Pd)	Type Formula	Kinship				
	10-M-0009-6	10-M-1009-1	PC	FS	HS	FC	UR							
D3S1358	15 / 16	15 / 16	15 (0.4)	16 (0.321)	0 0	0 0	abab	$\Phi_2 + 0.5\Phi_1(Pa+Pb) + 2\Phi_0 PaPb$	0.361	0.494	0.309	0.283	0.257	
TH01	9 / 9.3	8 / 9	9 (0.464)	9.3 (0.064)	8 (0.048)	0 0	abac	$0.5\Phi_1 Pc + 2\Phi_0 PaPc$	0.024	0.023	0.034	0.039	0.045	
D21S11	30.2 / 31.2	30.2 / 30.2	30.2 (0.009)	31.2 (0.045)	0 0	0 0	abaa	$\Phi_1 Pb + 2\Phi_0 PaPb$	0.045	0.023	0.023	0.012	0.001	
D18S51	17 / 19	14 / 19	19 (0.033)	17 (0.076)	14 (0.188)	0 0	abac	$0.5\Phi_1 Pc + 2\Phi_0 PaPc$	0.094	0.050	0.053	0.033	0.012	
D5S818	11 / 12	11 / 11	11 (0.312)	12 (0.242)	0 0	0 0	abaa	$\Phi_1 Pb + 2\Phi_0 PaPb$	0.242	0.159	0.197	0.174	0.151	
Penta_E	5 / 13	5 / 7	5 (0.067)	13 (0.045)	7 (0.015)	0 0	abac	$0.5\Phi_1 Pc + 2\Phi_0 PaPc$	0.007	0.004	0.005	0.003	0.002	
D13S317	12 / 13	12 / 12	12 (0.194)	13 (0.021)	0 0	0 0	abaa	$\Phi_1 Pb + 2\Phi_0 PaPb$	0.021	0.013	0.015	0.011	0.008	
D7S820	10 / 11	10 / 10	10 (0.215)	11 (0.373)	0 0	0 0	abaa	$\Phi_1 Pb + 2\Phi_0 PaPb$	0.373	0.227	0.267	0.214	0.160	
D16S539	9 / 11	9 / 12	9 (0.294)	11 (0.255)	12 (0.194)	0 0	abac	$0.5\Phi_1 Pc + 2\Phi_0 PaPc$	0.097	0.077	0.106	0.110	0.114	
CSF1PO	10 / 12	10 / 13	10 (0.239)	12 (0.394)	13 (0.079)	0 0	abac	$0.5\Phi_1 Pc + 2\Phi_0 PaPc$	0.040	0.029	0.039	0.038	0.038	
Penta_D	12 / 15	12 / 12	12 (0.145)	15 (0.018)	0 0	0 0	abaa	$\Phi_1 Pb + 2\Phi_0 PaPb$	0.018	0.010	0.012	0.008	0.005	
VWA	17 / 18	17 / 18	17 (0.264)	18 (0.194)	0 0	0 0	abab	$\Phi_2 + 0.5\Phi_1(Pa+Pb) + 2\Phi_0 PaPb$	0.229	0.390	0.166	0.134	0.102	
D8S1179	13 / 14	13 / 14	13 (0.224)	14 (0.152)	0 0	0 0	abab	$\Phi_2 + 0.5\Phi_1(Pa+Pb) + 2\Phi_0 PaPb$	0.188	0.361	0.128	0.098	0.068	
TPOX	8 / 11	8 / 8	8 (0.448)	11 (0.333)	0 0	0 0	abaa	$\Phi_1 Pb + 2\Phi_0 PaPb$	0.333	0.241	0.316	0.307	0.298	
FGA	20 / 25	20 / 25	20 (0.055)	25 (0.073)	0 0	0 0	abab	$\Phi_2 + 0.5\Phi_1(Pa+Pb) + 2\Phi_0 PaPb$	0.064	0.284	0.036	0.022	0.008	
	Parent-Child	Full-sib		Half-sib		First Cousin		Unrelated						
Multiplication	3.1161769679510774E-17	5.276851409356706E-18		6.058138702941143E-19		1.5928738739881584E-20		1.9213991028604402E-23						
Index(VS Unrelated)	1621827.0130926669	274635.88389840047		31529.829976095145		829.0177046594863		1.0						
Probability	99.99993834118094 %	99.999635882848 %		99.99682850086484 %		99.87952064222411 %		50.0 %						

도면12



도면13

