



저작자표시-비영리 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

김 종 진 교수지도

석사학위 청구논문

실종아동 DNA 데이터베이스의 검색
및 친자확인 시스템 개발

2012

성신여자대학교 대학원

생물학과

김영은

실종아동 DNA 데이터베이스의 검색 및 친자확인 시스템 개발

김 종 진 교수지도

이 논문을 석사학위논문으로 제출함

2012년 5월

성신여자대학교 대학원

생물학과

김영은

인 준 서

김영은의 석사학위 논문으로 인준함.

심사위원_____인

심사위원_____인

심사위원_____인

성신여자대학교 대학원

논문개요

인간 유전체 내 개인식별이 가능한 STR (short tandem repeats) 마커가 발견됨에 따라 유전자에 의한 친자확인이 가능하게 되었다. 또한 미토콘드리아 DNA를 이용한 모계혈통과 Y-STR을 이용한 부계혈통의 가족관계 검사가 가능하다. 최근에는 X-STR을 이용하여 부녀관계를 검사하고 있다.

2005년 『실종아동등의 보호 및 지원에 관한 법률』이 시행됨에 따라 유전자검사결과에 대한 데이터베이스를 구축하여 운영하여 왔다. 데이터베이스에 수록되어 있는 실종자와 실종자가족에 대하여 가족관계확인을 한다. 멘델의 유전 법칙에 따라 자녀의 DNA는 모계와 부계로부터 각각 한 세트의 대립유전자를 받으므로 상 염색체 STR의 대립유전자 공유여부를 검색하는 시스템으로 친자관계확인을 할 수 있다. 기존에 이용되어 온 검색시스템은 그룹 간 비교 기능이 없으며, 돌연변이 검색 기능도 없어 친자확인검색에 많은 시간이 소용되는 단점이 있었다. 따라서 본 시스템개발연구에서는 이러한 부분을 개선하는데 주안점을 두었다. 본 연구에서 개발한 실종아동 DNA 데이터베이스 시스템은 실종아동군, 변사자군, 실종아동 보호자군, 변사자보호자군을 주요 4그룹으로 지정하고 그 외 N.D.군, 대량사망사고군, 대량사망보호자군을 추가하였다. 본 시스템은 실종아동군과 보호자군을 1:1로 비교하는 동일성 및 부모자식관계 검사 기능, 비교 분석을 원하는 데이터와 데이터베이스 내에 있는

모든 프로필을 비교하는 동일성 및 부모자식관계 검색 기능, 그룹 간 비교 검색할 수 있는 일괄검색 기능을 갖추었다. 검색결과는 대립유전자의 공유율과 친부지수(Paternity index)의 값으로 나타나고, 대립유전자의 돌연변이를 고려한 허용오차에 따라 검색결과 후보군(Candidates)의 범위를 결정지음으로써 실제 가족관계를 추정하는데 소요되는 시간과 노동력을 최소화하였다. 본 시스템은 돌연변이율을 고려한 검색시스템으로서 사용자의 편리성을 제고하여 친자확인 검사 및 검색업무의 효율성을 높일 것으로 기대된다. 또한 DNA 분석 결과는 최근의 과학적 통계계산법을 반영하고 유의성 정보를 제공함으로써 사회적 이슈가 되고 있는 가족관계확인을 신속하고 정확하게 찾아주는 데 기여할 것이다.

목 차

논문개요

I. 서론	1
II. 재료 및 방법	7
1. 데이터베이스 검색 시스템 개발 내용	7
2. 데이터 세트 구성	8
3. 친자확인 DNA정보 검색시스템 사양 및 개발 운영환경	9
4. 통계 알고리즘	10
III. 결과	12
1. 데이터베이스	12
2. 등록	13
3. DNA 프로파일 검사.....	16
4. 검색	17
IV. 고찰	19

참고사이트

참고문헌

ABSTRACT

LIST OF TABLES

표 1. 경찰청 실종아동 접수 및 발견 현황	23
표 2. 미국의 친자확인 관련 기관 및 주요사업.....	24
표 3. 캐나다의 친자확인 관련 기관 및 주요사업.....	25
표 4. 우리나라의 친자확인 관련 기관 및 주요사업.....	25
표 5. 17개 A-STR 마커 정보.....	26
표 6. 16개 Y-STR 마커 정보	27
표 7. 8개 X-STR 마커 정보	28
표 8. 관계에 따른 일치 확률 공식	29
표 9. Trio 검사 시 PI값 계산.....	30
표 10. Duo 검사 시 PI값 계산.....	30
표 11. 친부지수 검사 시 A-STR 대립 유전자 빈도.....	31
표 12. 등록메뉴의 하위메뉴 기능	34
표 13. 가족코드 번호	34

LIST OF FIGURES

그림 1. 친자확인 DNA 정보 검색 시스템 전체 개요	35
그림 2. 친자확인 DNA 정보 검색 시스템 주 화면	36
그림 3. 친자확인 등록.....	37
그림 4. 보호자 등록.....	38
그림 5. STR DNA Data 등록화면	39
그림 6. mtDNA 등록화면	40
그림 7. 동일성 검사 결과 화면	41
그림 8. 부모자식관계 검사 결과 화면	42
그림 9. 동일성 검색	43
그림 10. 부모자식관계 검색	44

I. 서론

2011년에 개정된 “실종아동등의 보호 및 지원에 관한 법률” 제2조에 의하면 ‘실종아동등’이라 함은 실종 당시 14세 미만인 아동 및 ‘장애인복지법’ 제2조의 장애인 중 지적장애인, 자폐성장애인 또는 정신장애인으로 약취·유인 또는 유기되거나 사고를 당하거나 가출하거나 길을 잃는 등의 사유로 인하여 보호자로부터 이탈된 아동등으로 정의하고 있다(법제처). 즉, 아동은 14세 미만으로 장애인은 성인을 포함하여 정의하고 있다. 따라서 본 논문에서는 단순히 길을 잃어버린 아동을 의미하는 협의적 개념인 미아라는 용어보다는 포괄적인 의미인 친자확인 대상 아동(missing child)이라는 용어를 사용하고자 한다(김종우, 2006).

우리나라에서 친자확인 사업을 조직적으로 전개하기 시작한 시기는 1986년이다. 친자확인에 대한 제도적·법률적 대책이 국가적 차원에서 없었기 때문에 친자확인 대상 아동 부모들이 자체적으로 모임을 결성하여 전국의 아동복지시설에 방문 친자확인에 직접 나서게 되면서 사회적으로 이슈화 되었다. 이로 인해 친자확인 사업에 대한 사회적 인식이 확대되었고 그에 따라 정부의 대책이 필요하게 되었다. 이에 정부에서는 사회복지법인 한국복지재단 내에 어린이 찾아주기 종합센터를 설치하여 친자확인 사업을 시작하게 되었다. 또한 근거를 마련하고자 2005년 5월 31일에 “실종아동등의 보호 및 지원에 관

한 법률”을 제정하고 친자확인전문기관을 설치하였고 이를 보건복지부와 경찰청이 공동으로 운영하게 되었다. 기존의 어린이 찾아주기 종합센터를 운영하고 있던 한국복지재단은 친자확인전문기관의 운영을 보건복지부로부터 위탁 받아, 경찰청과 공조하여 친자확인 사업을 보다 체계적이고 조직적으로 전개함으로써 국가적 차원에서 친자확인 사업을 주관하고 관리할 수 있게 되었다 (김종우, 2006).

친자확인 사업의 절차는 다음과 같다. 친자확인 대상 아동이 발생하면 경찰청 실종아동 찾기 센터에 신고·접수되며 실종 1-2일 내에 발견되면 부모에게 인계된다. 48시간 이상 미발견 된 아동의 경우에는 장기 친자확인 대상 아동으로 분류되어 친자확인전문기관으로 관련 자료(대상 인적 사항 및 본인, 보호자 등의 후 무연고 아동 및 친자확인 가족 DNA시료)를 인계하여 후속절차를 진행한다. 친자확인 사업에서 가장 심각한 문제는 수개월 또는 몇 년이 지나도 부모를 찾지 못하는 아동들이 있다는 사실이다. 대부분의 장기 친자확인 대상 아동들은 실종 당시의 모습을 지니고 있지 않다. 아동의 얼굴이나 체형이 달라져 부모를 만나더라도 서로 알아보지 못하는 경우가 많기 때문에 기억에 의존하는 가족확인은 가장 큰 난관으로 작용하고 있다. 따라서 객관적이고 과학적인 접근의 필요성이 대두되었고 2005년 실종 아동 등의 보호 및 지원에 관한 법률이 제정됨에 따라 이 법률에 근거하여 2004년부터 보호시설에 수용되어 있던 무연고 아동 및 아동을 찾는 가족의 유전자를 채취하여 그 정보를

데이터베이스화하여 유전자 검사를 친자확인 사업에 적용함으로써 보다 과학적인 접근이 가능하게 되었다. 경찰청 통계에 따르면 2011년의 경우 보호자인 비율이 99.5%나 되지만 2012년에는 미발견 건수가 증가했다. 장애인의 경우 가정복지율이 비장애 아동에 비해 낮다. 표 1은 친자확인 대상 아동의 발생현황으로, 경찰청에서 발표한 자료를 토대로 아동과 장애인 그리고 보호자 인계 건수를 2006년부터 2012년 현재까지(2012년 3월 31일 기준) 정리한 것이다.

30억 개 이상의 염기쌍으로 이루어진 인간 게놈 중 단백질로 발현하는 유전자 암호화 영역(coding-region)은 전체 염기서열의 5 % 이내이다. 게놈의 대부분은 동일 염기서열이 반복되는 반복 DNA (repetitious DNA)와 기타 DNA를 포함하는 비유전자 DNA 영역 (non-coding region)이다. 비유전자 영역에서 반복염기서열은 개인별 길이 다형성(length polymorphism)을 특징으로 한다. 반복되는 염기수가 10-100 bp로 구성된 유전자 좌위를 variable number of tandem repeat (VNTR)라 하며, 염기수가 2-6 bp로 구성된 유전자 좌위를 short tandem repeat (STR: microsatellite)라 한다(Valdes et al., 1993). 인간 염색체는 아버지와 어머니로부터 각각의 haplotype이 유전되므로 길이 다형성도 고스란히 유전되어 염색체에 대립유전자(allele)의 형태로 존재하게 된다. 이와 같은 각 개체의 유전적 다형성(genetic polymorphism)을 개인 식별 및 친자감정 등에 이용하고 있다(Edwards et al., 1991). 유전적 다형성을 이용한 개인 식별에 관한 연구는 초기에는 주로 상염색체에 한정되어 이루어

져 왔다. 친자감정에 상염색체를 이용한 유전적 다형성을 적용하기 위해서는 많은 가족 구성원이 생존해 있어야 한다. 이에 대한 보완으로 부계 또는 모계만으로 유전하는 유전자 좌위를 추가적으로 확인하면 전체 유전자검사의 종합적 결론에 도움이 될 수 있다(Jobling et al., 1997; Forster et al., 1998; Horst et al., 1999; Decorte et al., 2000; Brenner, 2000). 부계확인을 위해서는 Y 염색체 상의 16개의 microsatellite 좌위에 대해 조사하며, 모계확인을 위해서는 미토콘드리아 DNA의 비유전자 서열 중 고도의 가변부위인 HV1(hypervariable 1)과 HV2 염기서열을 비교한다(Anderson et al., 1981; Greenberg et al., 1983). Y 염색체는 아버지로부터 아들에게로 그리고 손자에게로 그대로 전해진다. 부계 유전되는 Y 염색체로는 형제지간, 삼촌-조카, 조부-손자 간 등의 친족확인도 물론 부자간의 친자확인이 가능하다. 미토콘드리아 DNA 분석은 핵 대신 미토콘드리아로부터 추출한 DNA의 염기배열형태를 분석하는 방법으로 한 가계 안에서 동일 모계인 모든 사람들은 미토콘드리아 DNA가 같다. 따라서 부모가 없을 경우 모계가 동일한 자매간, 남매간, 이모-조카, 조모-손녀 또는 손자의 친족확인으로도 널리 쓰이고 있다(Holland et al., 1999).

스페인에서는 세계 최초로 실종자와 인간유해의 신원 확인을 위해 국영 프로그램인 A Missing Persons Genetic Identification Program (Phoenix program)을 만들어 사용하였다. 이 프로그램은 모계확인을 위해서 미토콘드리아 DNA의

염기서열만을 검색하게 되어 있다(Lorente et al., 2002; Bruce et al., 2005). 미국의 경우 1984년에 설립된 비영리기관인 Missing Children Awareness Foundation의 아동 ID 세트 프로그램은 DNA, 머리카락 및 개인정보, 아동의 최근 사진, 엄지손가락 지문, 신체 묘사, 치과 ID 차트를 포함하며 새로 나온 키트에서는 발자국도 포함시키고 있다. 표 2와 표 3과 표 4는 미국과 캐나다, 우리나라의 친자확인 등과 관련한 공·사적 기관과 그들이 수행하는 사업을 정리한 것이다(김성천, 2008). 우리나라의 경우에도 2005년 실종아동 등의 보호 및 지원에 관한 법률 제 11조에 의해 유전자검사 기관이 유전자검사 실시 결과를 데이터베이스로 구축 운영할 수 있게 됨에 따라 아이를 잃어버린 부모의 유전자(보호자군)와 부모를 잃어버린 아이의 유전자(아동군)의 데이터를 지속적으로 축적하여 왔다. 그러나 데이터가 증가됨에 따라 추가 발생하는 친자확인의 신속한 검색 및 분석에 애로를 겪고 있어서 기 구축된 친자확인 대상 아동-부모 데이터베이스의 유전자형 분석결과의 체계적 관리가 필요하게 되었다.

따라서 본 논문에서는 기 구축된 친자확인 대상 아동-부모 데이터베이스의 재구성을 위해 친자확인 대상 아동의 특징정보(성별, 연령) 및 친자확인 대상 아동 및 부모의 Autosomal-STR (A-STR), Y-STR, X-STR, mtDNA 관리를 위한 DNA 데이터베이스 시스템을 구축하고, DNA 실험장비로부터 데이터베이스까지 실제 데이터를 다루는 번거로움을 최소화함으로써 친자확인 검색 업무의 효율

성을 높일 것으로 기대된다. 또한 DNA 분석결과는 최근의 과학적 통계계산법을 반영하고 유의성 정보를 제공함으로써 사회적 이슈가 되고 있는 친자확인을 신속하고 정확하게 수행하는 데 기여할 것이다.

II. 재료 및 방법

1. 데이터베이스 검색 시스템 개발 내용

새롭게 개발하는 실종아동 DNA 데이터베이스 시스템에서는 mtDNA 를 포함하여 A-STR 실험데이터와 Y-STR 과 X-STR 데이터를 추가 실험한 자료도 관리할 수 있도록 시스템의 기능을 확장하고자 하였다. 또한 시스템은 친자확인전문기관으로부터 의뢰되는 친자확인 대상 아동 정보(성별, 나이)와 부모정보를 입력하여 1:1 matching 검사뿐만 아니라 1:n matching 검사, n:n matching 검사 등이 가능하도록 검사 및 검색 기능을 향상시켰다. 그리고 시료중복검사 및 오염검사가 가능하도록 하여 데이터베이스의 품질관리 기능을 추가하였다(그림 1).

실종아동과 보호자군의 DNA 데이터베이스 자료는 30,000여건에 달하고 있기 때문에, 돌연변이를 고려하지 않고 1좌위 만 상이해도 matching에서 제외시킬 경우 가족관계확인이 누락될 수 있는 가능성이 있다. 따라서 이러한 문제점을 개선하고자 돌연변이로 인해 나타날 수 있는 1회 반복 차이의 1좌위 상이한 후보군도 모두 찾을 수 있도록 시스템을 개선하였다.

2. 데이터 세트 구성

2.1 STR dataset

STR 데이터는 A-STR, Y-STR, X-STR 로 구성되어 있으며, 현재 A-STR, Y-STR 그리고 X-STR 에는 전부 25,619 건의 STR 데이터가 구축되어 있다. A-STR 데이터에 사용된 좌위는 성염색체를 구분할 수 있는 Amelogenin 을 포함하여, TH01, TPOX, CSF1PO, D3S1358, vWA, FGA, D5S818, D13S317, D7S820, D16S539, D8S1179, D21S11 그리고 D18S51 의 13 개 Combined DNA Index System(CODIS, USA) 을 기본으로 수록하고, AmpFISTR Identifiler™ kit (Applied Biosystems, USA) 좌위인 D2S1338, D19S433 과 PowerPlex 16 (promega, USA)의 Penta E, Penta D 를 추가적으로 수록하여 총 17 개 좌위를 수록하였다(표 5). 또한 Y-STR 좌위는 DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385, DYS393, DYS391, DYS439, DYS635, DYS392, DYS437, DYS438, DYS448 그리고 GATAH4.1 총 16 개 좌위로 구성하였다(표 6)(John M, 2005). X-STR 은 Mentype® Argus X-8 kit (Bio type, Germany)를 사용해서 DXS7132, DXS7423, DXS8378, DXS10074, DXS10101, DXS10134, DXS10135, HPRTB 로 구성하였다(표 7).

2.2 Mitochondrial DNA (mtDNA)

개인식별을 위해서는 다형성이 높은 부위를 검사대상으로 하여야 하고 mtDNA에서는 D-loop가 이에 해당한다. 이 부위는 mtDNA가 복제할 때 이중나선이 먼저 풀려지는 부위이기 때문에 D-loop라고 명명되어졌으며, 이곳에는 여러 프로모터 부위가 있어 mtDNA의 복제와 전사에 특정 역할을 하는 것으로 여겨져 조절지역(Control region)이라고 부르기도 한다(John M. Butler, 2005). 흔히 앤더슨 표준염기서열 16024~577를 D-loop라고 하며 과변이 제1영역(Hypervariable region I; HV 1), 과변이 제2영역(HV 2) 및 제3영역(HV 3)으로 구성되어 있다. 본 시스템에서 설정된 과변이 제 1영역은 미토콘드리아 유전체 (genome)의 16024~16365 bp 지역에 위치하고, 제 2영역은 73-340 bp에, 제 3영역은 438~574 bp에 위치하고 있다.

3. 친자확인 DNA정보 검색시스템 사양 및 개발 운영환경

친자확인 DNA 정보 검색시스템은 검색시스템, 분석 데스크탑, 분석 노트북 총 3대의 하드웨어로 구성되어 있다. 검색시스템은 Intel 쿼드코어 E5504 Xeon CPU (2.0GHz) 프로세서와 1TB의 저장 하드 그리고 24GB 메모리를 탑재하였고, Operation System (OS)는 Microsoft window 서버 2008을 바탕으로 개발되었다. 분석 시스템은 Intel® Core™프로세서 i7-860 (2.80 Hz), 500 GB 하드 용량, 4 GB의 메모리를 가진 검색시스템과는 별도의 데스크탑에서 개발되

어 분석을 수행하게 된다. 본 시스템은 Apache 웹서버 환경에서 PostgreSQL DBMS (Database Management System)를 바탕으로 Python 언어를 이용하여 개발되었다.

4. 통계 알고리즘

친자확인 DNA 정보 검색시스템은 친자확인 대상 아동을 찾고자 하는 친부/친모의 대립유전자 DNA 프로파일과 친자확인 대상 아동의 DNA 프로파일을 비교하여 각각의 일치확률(Match Probability)과 친부지수(Paternity Index) 계산을 통해 통계적인 결과를 얻을 수 있다.

4.1 일치 확률 공식(Match Probability formula)

일치확률(match provability)은 각 좌위에서 두 개의 대립유전자가 공유될 때 확률을 계산하는 가장 보편화된 공식이다. 두 대립유전자가 동형접합자일 때와 이형접합자일 때의 계산식은 다르다(표 8).

4.2 친부지수(Paternity Index)

자식과 부모의 프로파일을 모두 알고 있는 경우에 검사하는 Trio 검사방법(표 9)과 부모 중 한 명을 아는 경우(Duo 검사)은 다르다(표 10). 친부지수 감정을 위해서는 추정부(AF: alleged father)와 어머니(M: mother), 그리고 아이(C: child)에 대한 유전자 검사 결과가 필요하다. 본 시스템에 적용된 가족관계 확률을 구하기 위해서는 A-STR 마커 각각의 allele frequency를 곱하기 법

칙으로 계산하여 likelihood ratio를 확정한다. 각각의 가족관계 별로 계산된 확률 중 가장 높은 수치의 결과를 확인하여 가족관계 여부를 결정한다. 본 시스템에 적용되는 각각의 빈도값은 표 11에 정리하였다(Cho NS et al., 2005).

Ⅲ. 결 과

친자확인 DNA정보 검색시스템은 트리 구조의 7개의 메뉴로 되어 있다. 데이터베이스 메뉴는 전체현황을 관리하고 확인할 수 있고, 데이터베이스 하위 주요 4그룹인 실종아동군, 보호자군을 비롯하여 변사자군과 변사자보호군으로 대상군들을 나누어서 등록할 수 있게 하였다. 등록 메뉴는 신원 및 정황정보와 STR 데이터를 개별 혹은 일괄적으로 입력 가능하게 하였고, 주요 4그룹을 나누어 입력 가능하게 하였다. 검사 메뉴는 동일 프로필 여부를 확인하는 동일성 검사 및 부모자식 관계 검사 기능을 가능하게 하였다.

검색 메뉴로는 비교 분석을 원하는 검체를 데이터베이스 내에 있는 모든 검체와 비교하는 검색, 다수의 검체를 다수의 검체와 비교하는 일괄검색이 가능하며, 분석결과에서는 시스템에서 수행한 분석 결과들을 목록형식으로 볼 수 있다. 그 외 대량사망사고 발생 시 희생자 분석을 수행하는 대량사망사고 메뉴가 있다(그림 2).

1. 데이터베이스

본 시스템에서 데이터베이스 메뉴는 전체현황을 나타내며 『실종아동현황』, 『보호자현황』, 『변사자 현황』, 『변사자 보호자 현황』으로 구성되어 있다. 전체현황에서는 등록된 검체들의 정보의 변경이나 삭제 등의 관리를 할

수 있으며 실종아동현황, 보호자 현황, 변사자 현황, 변사자 보호자 현황에서는 입력된 레코드의 개수와 등록된 정보 현황의 개별적인 확인만 가능하다.

2. 등록

등록 메뉴에서는 검체를 새로 등록하고 STR결과를 등록할 수 있다. 『등록』 메뉴는 4개의 하위 메뉴로 구성되어 있으며 각 하위 메뉴는 다음과 같은 기능을 가진다(표 12). 입력 방법은 신원정보를 개별 또는 일괄 입력 후 DNA 데이터(A-STR, Y-STR, X-STR, mtDNA)를 등록하는 방식이다.

2.1 의뢰데이터 개별입력 및 일괄입력

의뢰 데이터를 한 건씩 수동으로 입력할 수 있도록 한 개별입력과 두 건 이상의 정황정보 데이터를 함께 입력 할 수 있는 일괄입력으로 나누었으며 각각의 하위메뉴 구성은 동일하다.

2.1.1 친자확인 대상 아동 등록

메뉴에서 『등록』 - 『의뢰데이터 개별입력(의뢰데이터 일괄입력)』 - 『친자확인 대상 아동 또는 변사자』를 통해서 친자확인 대상 아동 및 변사자의 데이터를 등록할 수 있다(그림3). 각 검체에는 검체번호를 부여하여 등록하는데, 기관에서 부여해서 보내 오거나 시스템 관리자가 직접 부여하여 입력한다. 검체번호는 C(Child), 의뢰 연도, 성별(F: 여성, M: 남성)로

구분하여 다음과 같이 C-yy000000-F/M의 형식으로 표기한다. 성별과 검체번호는 필수 입력사항이며 이외에도 검체를 의뢰한 곳의 의뢰번호, 요약설명, 생년월일 등의 정보를 입력할 수 있다. 변사자의 입력 항목들은 친자확인 대상 아동과 동일하다. 친자확인 대상 아동에 변사자가 포함될 수 있기 때문이다.

2.1.2 보호자, 변사자 보호자 등록

보호자 데이터의 등록은 친자확인과 마찬가지로 메뉴에서 『등록』 - 『의뢰데이터 개별입력(의뢰데이터 일괄입력)』 - 『보호자(변사자 보호자)』를 통해 등록할 수 있다(그림 4). 식별코드는 친자확인 대상 아동의 경우와 같이 F (Family), 의뢰년도, 가족코드 번호를 표기하여 다음과 같은 형식 F-yy000000-1A-M/F으로 나타내며 가족코드는 표 13에서 정리하였다. 그리고 친자확인 대상 아동과 마찬가지로 의뢰번호, 요약, 생년을 입력하고 찾고자 하는 아동의 실종시기를 추가로 입력할 수 있다. 변사자 보호자 항목은 변사자 항목이 추가됨으로써 생성된 것이다.

2.2 DNA 일괄입력

2.2.1 A-STR, Y-STR, X-STR 등록

본 시스템에 적용된 A-STR 데이터에 사용된 좌위는 성염색체를 구분할 수 있는 Amelogenin을 포함하여, TH01, TPOX, CSF1PO, D3S1358, vWA, FGA, D5S818, D13S317, D7S820, D16S539, D8S1179, D21S11 그리고 D18S51의 13개

Combined DNA Index System (CODIS, USA)을 기본으로 하여 총 17개 좌위를 수록하였다(표 3). 그림 5는 A-STR 등록화면이다. Y-염색체에도 상당수의 STR이 보고되어 있으며 CODIS의 표준 STR보다 대립유전자의 수는 적지만 여러 곳의 STR 프로파일은 상당한 식별력을 가진다. Y-STR 좌위는 DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385, DYS393, DYS391, DYS439, DYS635, DYS392, DYS437, DYS438, DYS448 그리고 GATAH4.1의 총 16개 좌위로 구성되어 있다(표 6)(John M., 2005). X-STR좌위에는 DXS7132, DXS7423, DXS8378, DXS10074, DXS10101, DXS10134, DXS10135, HPRTB 등이 있으며 자매간 혹은 친할머니-친손녀, 이복자매간 혈연확인이 가능하다(표 7).

2.2.2 mitochondrial DNA (mtDNA) 등록

본 시스템에서 『mtDNA 등록』은 STR 마커와 다른 방식으로 등록된다. 앤더슨 표준염기서열의 16021~16380 bp의 과변이 제1영역(HV 1), 31~408 bp의 제2영역(HV 2)이 개인식별에 이용되며, 변이 영역의 위치와 염기를 공백으로 구분하여 등록한다(예: 16024t 16330A)(그림 6).

3. DNA 프로파일 검사

DNA 프로파일 검사는 비교를 원하는 검체번호를 입력함으로써 시작되며, 두 개체간의 마커 정보를 비교하여 일치 여부를 보여준다. 동일성 검사 및 부모자식관계 검사로 구성되었다.

3.1 동일성 검사

각 마커의 대립유전자의 일치 여부로 판단하며 각 마커의 2개의 대립유전자가 모두 일치하여야 일치로 판정된다. 검체번호를 넣고, 옵션조건을 선택한 후에 검색을 누르면 결과 화면을 볼 수 있으며, 상단에는 일치여부, 식별지수, 불일치수, 허용오차가 표시된다. 옵션조건 of 불일치수는 불일치하는 마커의 개수를 허용해주는 옵션이며 허용오차는 불일치수에서 허용할 오차 범위를 지정하는 옵션이다. 만일 불일치일 경우라도 불일치수 설정 범위 내에서 허용오차 범위 내로 들어오면 허용이 된다(그림 7).

3.2 부모자식관계 검사

부모자식관계검사는 A라는 검체와 B라는 검체가 친자관계가 있는지를 확인하는 검사이다. 즉 각각의 마커의 대립유전자중에 공유하는 데이터가 있는지 여부로 판단한다. 검체번호를 넣고 옵션조건을 선택한 후에 검색을 누르면 결과 화면을 볼 수 있는데, 상단에는 일치여부, 식별지수, 허용오차가 표시된다. 부모자식관계 검사는 추가로 아래아한글로 다운 받을 수 있으며 분석결과를 저장할 수 있다(그림 8).

4. 검색

검색은 1:n 분석을 수행할 수 있는 메뉴이다. 1개의 검체에 대하여 “동일성 검색” 또는 “부모자식관계검색” 을 특정 그룹군 또는 전체에

대하여 수행할 수 있다. 비교 분석을 원하는 검체번호를 입력한 뒤에 검색대상 그룹 구분 선택 버튼을 누르면 데이터베이스 내에 있는 모든 검체들의 정보와 비교 분석하여 일치하는 검체의 목록을 출력한다.

4.1 동일성 검색

본 시스템에서 『검색』 - 『동일성 검색』을 선택하면 그림 9의 화면이 나타난다. 검체번호를 넣고 검색조건을 선택한 후 검사 버튼을 누르면 동일성 검사가 진행된다. 검사가 완료되면 일치된 검체 목록이 나오며, 『검체번호』를 클릭하면 그 검체의 상세 정보를 확인할 수 있으며 검사결과에 『일치』를 클릭하면 두 검체 간의 각 좌위 별로 일치 여부를 확인할 수 있다(그림 9).

4.2 부모자식관계 검색

본 시스템에서 『검색』 - 『부모자식관계 검색』을 선택하면 그림 10의 화면이 나타난다. 검체번호를 넣고 검색 조건을 선택 한 후 검색 버튼을 누르면 부모자식관계 검색이 진행된다. 옵션조건인 부분검색허용은 두 대상의 프로필이 가지고 있는 STR 마커 수가 다를 경우, 서로 공유하는 마커 만을 대상으로 검사를 시행한다. 예를 들어, 첫 번째 검체에 A, B, C 마커의 실험데이터가 있고 두 번째 검체에 B, C, D 마커의 실험데이터가 있을 시, B, C만을 비교해서 검사를 한다. 검사 대상 마커수가 17개이지만 옵션 조건 불일치수를 1로 지정하면 17좌위 중 16좌위가 공유될 경우 16/17로 표시된다.

부분검색허용 시에는 검사 대상 마커 수가 더 줄어들 수 있기 때문에 공유마커 비율은 작아지게 된다. 『검색 번호』를 클릭하면 그 검체의 상세 정보를 확인할 수 있으며 검사결과의 『일치』를 클릭하면 두 검체 간의 각 좌위 별로 일치 여부를 확인할 수 있다(그림 10).

IV. 고찰

2005년 “실종아동등의 보호 및 지원에 관한 법률”이 제정되고 동법 제 11조에 의해 유전자분석 결과를 데이터베이스로 구축 운영하여 왔다. 초창기의 데이터베이스 시스템은 실종 아동과 보호자의 가족관계를 확인을 위한 단순 검색 기능을 가지고 있었으며, STR의 돌연변이에 의한 대립유전자의 차이를 고려하지 않아 17개 모든 좌위에서 공유하는 경우에만 해서 친자관계의 후보군으로 선정하도록 구성되어 있었다. 그러나 STR은 돌연변이가 빈번하게 일어나며 돌연변이가 일어난 후보군의 경우는 검색 과정에서 배제될 수 밖에 없었다.

STR에서 돌연변이가 일어난 경우는 대부분 1개 좌위의 돌연변이다. 돌연변이 빈도는 유전자 좌위에 따라 다르며 인종간에 차이가 있기 때문에 친자확인용 데이터베이스에는 돌연변이를 감안한 검색 기능이 있어야 한다. 따라서 이번에 개발한 데이터베이스 시스템에는 검색옵션을 설정하여 불일치수와 허용오차범위를 설정하여 검색할 수 있도록 구현하였다. 그러나 이러한 돌연변이를 고려한 검색 기능의 문제점은 후보군이 너무 많이 나오는 경우가 종종 있어 추가 유전자분석을 하는 애로 사항이 있다는 것이다. 이러한 단점은 STR 마커의 수를 더 늘려 후보군을 줄임으로써 어느 정도 해결이 가능하리라 본다. 친자확인을 위한 검색과정은 A-STR, Y-STR, X-STR, mtDNA의 분석 결과를 matching하여 일치하는 동일 대상을 찾아서 액세스

구현되는 kinship analysis에서 처리하여 판정하였다. 그러나 이러한 방법으로 나온 실험 데이터 분석은 다시 정황정보(성별 및 나이)로 신원을 확인해야 하는 문제점이 있으므로 이를 해결하여야 한다. 실종아동 DNA 데이터베이스의 검색 및 친자확인 시스템을 적용하면 기존의 육안으로 확인하던 것보다 신속하고 정확하게 신원확인을 할 수 있을 것이며, 또한 신원정보 데이터와 실험 데이터의 일원화로 업무 효율성을 증대시킬 수 있을 것으로 판단된다.

본 연구에서는 실종아동군과 보호자군의 신원정보를 개별적으로 혹은 일괄적으로 등록이 가능하도록 하였다. 또한 실험데이터(A-STR, X-STR, mtDNA, Y-STR)도 개별등록과 일괄등록이 가능하며 상호검증을 거친 데이터를 입력하도록 함으로써 실험 과정 동안 오염되었는지를 지속적으로 확인할 수 있게 하였다. 등록된 검체의 데이터베이스를 바탕으로 검색도구를 이용하여 실험데이터 상호간을 비교 검색하여 부모자식관계나 가족관계가 확인이 되면 통계적 계산을 통해 부모-자식 관계 확률이 높은 대상을 식별한다. 그리고 원본데이터가 입력되고 A-STR 데이터가 입력되면 데이터 상세정보 화면에서 크로마토그램도 볼 수 있게 하였다. 크로마토그램은 원본데이터를 보정하여 구현하였다.

현재 시스템에 적용된 A-STR 마커 17개 중에서 multiplex STR profile에서 Tri-allelic pattern (Trisome)이 관찰될 수 있는데 주로 이러한 형태는 TPOX,

D13S317, D18S51, D21S11, Penta E에서 발생빈도가 높다. 그러나 실종아동 DNA 데이터베이스의 검색 및 친자확인 시스템에서는 기본적으로 2좌위 만 입력되도록 구현되어 있어 앞으로 보완이 요구된다. 각각의 부모자식관계 별로 계산된 확률 중 가장 높은 수치의 결과를 확인하여 가족관계 여부를 결정하는데 검사 옵션으로 불일치 수, 허용오차, 부분검색, C-Stretch 영역 제외 항목을 넣었다. 아버지가 딸을 찾는 경우 mtDNA나 Y-STR은 사용할 수 없고 A-STR 분석결과에만 의존해야 하므로 부분 검색이 필요하다. C-Stretch 영역은 variation이 매우 심한 영역으로써 mtDNA의 random matching 시 분석 결과에 영향을 미치기 때문에 필요에 따라 옵션으로 이용 가능하도록 프로그램을 구성하였으며, 모든 비교검색 기능에서 세부적인 옵션을 선택하여 능률적으로 운영되도록 하였다. 또한 테러 및 천재지변 등의 발생시 DNA 식별을 할 수 있도록 메뉴에 대량사망사고 항목을 넣어 국가적 긴급사항 시에 유용하게 이용되도록 하였다.

세계적으로 가장 먼저 실종자와 변사자의 신원확인을 위해 국영 프로그램(A Missing Persons Genetic Identification Program (Phoenix program))을 운영하고 있는 스페인은 모계확인을 위해 미토콘드리아 DNA 의 염기서열만을 검색할 수 있는 Reference Database (RD)을 운영하고(Lorente et al., 2002). 네덜란드의 경우는 2007 년 The Netherlands Forensic Institute (NFI)에서 신원확인 DNA database 인 Bonaparte software 를 개발하였다. Bonaparte

software 에서는 A-STR 과, 부계확인을 위한 Y-STR 과, 모계확인을 위한 mtDNA 를 수행할 수 있게 설계되었다(van Dongen ., 2011). 부계확인을 위해 본 시스템에는 X-STR 과 Y-STR 을 검색할 수 있게 하였으며, 부녀관계 및 친할머니-친손녀, 자매간, 이복자매간을 찾을 수 있도록 X-STR 을 추가하였다. 이러한 프로그램은 각국의 역사적 사회적 상황에 맞는 항목을 두어 운영될 필요가 있다. 우리나라는 역사적으로 8·15 광복과 한국전쟁을 거치는 동안 많은 이산가족이 생겨났다. 8·15 광복 후 삼팔선으로 길이 막혀 귀환이 불가능하게 된 경우와 한국전쟁 중 월남했거나 강제 납북된 경우이다. 그 수는 남북한을 합해 약 1,000 만 명이며, 중국에 약 200 만 명의 교포, 러시아에 약 40 만 명의 교포가 살고 있는 것으로 추정된다. 분단이 낳은 우리나라의 이산가족 문제는 인적 교류가 활발하게 된다면 다양한 형태의 인척관계 확인 문제로 대두될 것이다. 또한 2010 년 법무부 자료에 의하면 국제 결혼으로 인해 발생한 이민자는 136,556 명으로서 이로 인해 양산된 다문화가정이 우리 사회 10 만을 차지하고 있다. 법의학적 이용은 인종 고유의 자료를 필요로 하므로 그에 따른 지속적인 연구가 필요하며, 시스템의 개선을 위해 프로그램의 업그레이드가 필요할 것으로 판단된다.

표 1. 경찰청 실종아동 접수 및 발견 현황

연도/ 사례	발생 건수	아동		발생건수	장애인	
		발견건수 (보호자인계)	미발견 건수		발견건수 (보호자인계)	미발견 건수
2006	7071	7058 [99.8%]	13	6872	6835 [99.4%]	37
2007	8615	8600 [99.8%]	15	7239	7206 [99.5%]	33
2008	9485	9469 [99.8%]	16	4864	4854 [99.7%]	10
2009	9257	9239 [99.8%]	18	5564	5536 [99.4%]	28
2010	10872	10824 [99.8%]	48	6699	6660 [99.4%]	39
2011	11425	11358 [99.5%]	67	7377	7278 [98.6%]	99
2012	2217	2136 [96.3%]	81	1507	1428 [94.7%]	79

(2012년 03월 말 기준)

표 2. 미국의 친자확인 관련 기관 및 주요 사업

기관	주요 사업	
NCMEC(The National Center for Missing & Exploited Children)	찾기	(1)Age progression: 얼굴변형프로그램 (2)AMBER(American Missing Broadcast Emergency Response) Alert: 실종아동 발생시 즉각적으로 방송하는 프로그램
	홍보	(1)실종아동 사진 홍보 (2)PICTURE THEM HOME 캠페인: 목적은 개인과 시민단체, 기업들로 하여금 실종아동을 찾는 데 노력을 기울이도록 하는 것
Office Of Juvenile Justice Delinquency Prevention(미 법무부 산하의 청소년 사법 및 비행방지실, OJJDP)	찾기	(1)미아 및 유괴아동 수사 원조 (2)아이들의 안전과 실종방지를 위한 출판물 제작 및 제공
	홍보	(1)실종아동 사진 홍보 (2)PICTURE THEM HOME 캠페인: 목적은 개인과 시민단체, 기업들로 하여금 실종아동을 찾는 데 노력을 기울이도록 하는 것
Missing Children Awareness foundation	예방	아동의 ID 세트 프로그램(DNA 샘플을 확보한 아동용 ID 를 무료 제공): 최근사진, 지문, 신체묘사, 치과 ID 차트, 개인정보 및 DNA 머리카락, 발자국
Child Seek Network	찾기	조사원조
	홍보	전단지과 배지 제공
FBI National Crime Information Center (NCIC)	찾기	(1)집행기관을 지원하기 위한 국가 정보 시스템을 구축, NCMEC 에게 관련 정보제공 (2)FBI 와 NCMEC 와의 관계: 인터넷을 이용한 어린이 관련범죄에 대해 공동으로 대응하기 위한 파트너십 체결
Vanished Children' s Alliance(VCA)	찾기	실종아동에 관한 공공인식광고를 TV 골든아워에 방송하기 위해 계획 및 제작하여 실종아동의 사진과 정보를 방송
AMECO(Association of Missing & Exploited Children' s Organizations): 미국과 캐나다가 함께 실시하고 있음	찾기	자원봉사자 자원 탐색과 경보: 경보가 울리거나 집행기관에 의해 AMBER 가 발동될 때, 자원봉사자 조정을 위해 자원봉사자를 발굴

표 3. 캐나다의 친자확인 관련 기관 및 주요 사업

기관		주요 사업
Child Find Ontario	찾기	(1)신고된 실종아동 부모에게 연락하여 상담: 사례조사 및 찾는 작업 수행. 미디어, 포스터 등으로 홍보
	예방	(1)실종예방교육: 부모, 학교, 경찰, 관련기관, 지역사회조직 등을 대상 (2)KidCheck ID Program: 아동의 인적 사항 및 지문을 날인 후 아동실종발생시 경찰관에게 제공 및 활용 (3)가출아동 대상 프로그램 (4)경찰관 대상 프로그램: 각 상황에 대한 올바른 대처방법 교육 (5)아동대상 예방지침 책갈피 제작·배포
MCSC(Missing Children Society of Canada)	찾기	(1)project: Reunite-경찰과 가출 및 납치된 아동의 가족들을 위해 개발된 MCSC의 종합 수색프로그램. 경찰과 긴밀한 협력을 이루고 있음
	홍보	(1)안전교육프로그램: 부모와 아동에게 안전에 관한 다양한 정보 제공 (2)대중인식프로그램: 캐나다 전방에 실종아동에 대한 인식을 증가시키기 위함 (3)자원봉사자 서비스 프로그램: 기금모금, 종합수색프로그램, 대중인식프로그램에 참여

표 4. 우리나라의 친자확인 관련 기관 및 주요 사업

기관		주요 사업
실종아동전문기관	찾기	(1)온라인 및 오프라인 매체를 활용하여 실종아동의 사진을 홍보 (2)유선, 내방, 인터넷 신고접수 경찰청과 자료공유 (3)홍보 영상 제작
	예방	(1)실종예방교육: 일반인 및 부모, 교사, 등을 대상 (2)아동대상 예방을 위한 부모용 지침서 제작·배포
경찰청 실종아동찾기 센터	찾기	(1)실종아동 신고 접수 (2)웹보 경보 시스템: 실종 및 유괴 발생시 TV, 라디오, 휴대전화 인터넷을 활용 신속한 상황 전파
	홍보	(1)아동 유괴예방 홍보

표 5. 17개 A-STR 마커 정보

Loci	Chromosomal Location	GenBank Accession	Repeat motif ISFG format	Allele Range	Physical Position
TH01	11p15.5	D00269	TCAT	3-14	Chr 11 2,156 Mb
TPOX	2p25.3	M68651	GAAT	4-16	Chr 2 1,436 Mb
CSF1PO	5q33.1	X14720	TAGA	5-16	Chr 5 149,484 Mb
D3S1358	3p21.31	NT_005997	[TCTG][TCTA]	8-21	Chr 3 45,543 Mb
vWA	12p13.31	M25858	[TCTG][TCTA]	10-25	Chr 12 19,826 Mb
FGA	4q31.3	M64982	CTTT	12.2-51.2	Chr 4 145,086 Mb
D5S818	5q23.2	G08446	AGAT	7-18	Chr 5 123,187 Mb
D13S317	13q31.1	G09017	TATC	5-16	Chr 13 80,52 Mb
D7S820	7q21.11	G08616	GATA	5-16	Chr 7 83,401 Mb
D16S539	16q24.1	G07925	GATA	5-16	Chr 16 86,168 Mb
D8S1179	8q24.13	G08710	[TCTA][TCTG]	7-20	Chr 8 125,863 Mb
D21S11	21q21.1	AP000433	Complex [TCTA][TCTG]	12-41.2	Chr 21 19,476 Mb
D18S51	18q21.33	L18333	AGAA	7-39.2	Chr 18 59,098 Mb
D2S1338	2q35-37.1	G08202	TGCC/TTCC	15-28	
D19S433	19q12-13.1	G08036	AAGG	9-17.2	
Penta E	15q	AC027004	AAAGA	5-24	
Penta D	21q	AP001752	AAAGA	2.2-17	

표 6. 16개 Y-STR 마커 정보

Marker Name	GenBank Accession	Repeat motif	Allele Range	Position (Mb)
DYS456	AC010106	AGAT	13-18	
DYS389I		TCTG TCCTA	10-15/24-34	13.92
DYS390		TCTA TCCTG	18-27	16.52
DYS389II		TCTG TCCTA	10-15/24-34	13.92
DYS458	AC010902	GAAA	13-20	
DYS19		TAGA	10-19	9.44
DYS385		GAAA	7-25	20.00, 20.04
DYS393		AGAT	8-16	3.04
DYS391		TCTA	6-13	13.41
DYS439		AGAT	8-15	13.83
DYS635	G42673	TSTA compound	19-25	
DYS392		TAT	7-18	21.78
DYS437		TCTA	13-17	13.78
DYS438		TTTTC	8-12	14.25
DYS448	AC025227	AGAGAT	20-26	
GATAH4.1	AC011751	TAGA	8-13	

표 7. 87] X-STR 마커 정보

Marker Name	GenBank Accession	Repeat motif	Allele Range	Position (Mb)
DXS7132	G08111	[TCTA]	9-17	13
DXS7423	AC109994	[TCCA] TCTG TCCT [TCCA]	1018	15
DXS8378	G08098	[CTAT]	7-15	12
DXS10074	AL356358	[AAGA]	4-21	14
DXS10101	AC004383	[AAAAG] GAAAGAAG [GAAA] A [GAAA] AAGA [AAAG ₅] AAAAAGAA [AAAG] AA	24.2-35	28.2
DXS10134	AL034384	[GAAA] GAGA [GAAA] AA [GAAA] GAGA [GAAA] GAGA [GACAGA] [GAAA] GTAA [GAAA] AAA [GAAA] AAA [GAAA]	29-44.3	35
DXS10135	AC003684	[AAGA] GAAAG [GAAA]	14.1-36.2	23
HPR1B	M26434	[CTAT]	8-17	12

표 8. 관계에 따른 일치 확률 공식

A) Homozygotes (A_iA_i)	
Relationship	Match probability formula
Full siblings	$\frac{(1 + P_i)^2 + (7 + 7P_i - 2P_i^2)\theta + (16 - 9P_i + P_i^2)\theta^2}{4(1 + \theta)(1 + 2\theta)}$
Parent and child	$\frac{2\theta + (1 - \theta)P_i}{(1 + \theta)}$
Half siblings	$\frac{[2\theta + (1 - \theta)P_i][2 + 4\theta + (1 - \theta)P_i]}{2(1 + \theta)(1 + 2\theta)}$
First cousins	$\frac{[2\theta + (1 - \theta)P_i][2 + 11\theta + 3(1 - \theta)P_i]}{4(1 + \theta)(1 + 2\theta)}$
Unrelated	$\frac{[2\theta + (1 - \theta)P_i][3\theta + (1 - \theta)P_i]}{4(1 + \theta)(1 + 2\theta)}$
B) Heterozygotes (A_iA_j)	
Relationship	Match probability formula
Full siblings	$\frac{(1 + P_i + P_j + 2P_iP_j) + (5 + 3P_i + 3P_j - 4P_iP_j)\theta + 2(4 - 2P_i - 2P_j + P_iP_j)\theta^2}{4(1 + \theta)(1 + 2\theta)}$
Parent and child	$\frac{2\theta + (1 - \theta)(P_i + P_j)}{2(1 + \theta)}$
Half siblings	$\frac{(P_i + P_j + 4P_iP_j) + (2 + 5P_i + 5P_j + 8P_iP_j)\theta + (8 - 6P_i - 6P_j + 4P_iP_j)\theta^2}{4(1 + \theta)(1 + 2\theta)}$
First cousins	$\frac{(P_i + P_j + 12P_iP_j) + (2 + 13P_i + 13P_j - 24P_iP_j)\theta + 2(8 - 7P_i - 7P_j + 6P_iP_j)\theta^2}{8(1 + \theta)(1 + 2\theta)}$
Unrelated	$\frac{[2\theta + (1 - \theta)P_i][\theta + (1 - \theta)P_j]}{(1 + \theta)(1 + 2\theta)}$

표 9. Trio 검사 시 PI 값 계산

C	M	AF	Condition	Numerator	Denominator	PI
AA	AA	AA		1	a	1/a
AA	AA	AB	$A \neq B$	1/2	a	1/2a
AA	AA	BC	$C \neq A, B$	0	a	0
AA	AB	AA	$A \neq B$	1/2	a/2	1/4a
AA	AB	AB	$A \neq B$	1/4	a/2	1/2a
AA	AB	BC	$C \neq A, B$	0	a/2	0
AB	AA	BB	$A \neq B$	1	B	1/b
AB	AA	AB	$A \neq B$	1/2	B	1/2b
AB	AA	CD	$C, D \neq B, A$	0	B	0
AB	AB	AA	$A \neq B$	1/2	(a+b)/2	1/(a+b)
AB	AB	AB	$A \neq B$	1/2	(a+b)/2	1/(a+b)
AB	AB	BC	$C \neq A, B$	1/4	(a+b)/2	1/2(a+b)
AB	AB	CD	$C, D \neq B, A$	0	(a+b)/2	0
AB	AC	BB	$C \neq A, B$	1/2	b/2	1/b
AB	AC	BD	$D \neq A, B$	1/4	b/2	1/2b
AB	AC	CD	$C, D \neq B, A$	0	b/2	0

표 10. Duo 검사 시 PI 값 계산

C	AF or AM	Numerator	Denominator	PI
AB	AC	0.5b	2ab	0.25/a
AB	AB	0.5(a+b)	2ab	(a+b)/4ab
AB	A	b	2ab	0.5/a
A	AC	0.5a	a ²	0.5/a
A	A	a	a ²	1/a

표 11. 친부지수 검사 시 A-STR 대립 유전자 빈도

Allele	D6S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
5			0.0010					0.2660							
6						0.1570									
7			0.0020	0.0020		0.2350	0.0050					0.0020			
8			0.1310			0.0500	0.2850					0.5170			
9	0.002		0.0470	0.0370		0.4940	0.1410					0.1140			
9.1			0.0040			0.045									
9.2									0.0010						
9.3						0.0450									
10	0.1260		0.2020	0.2310		0.0170	0.1340	0.1620				0.0240	0.0010		
10.1			0.0010												
11	0.1000		0.3210	0.2400		0.0020	0.2190	0.2360		0.0040		0.3070	0.0070		
11.2										0.0020					
12	0.1430		0.2600	0.3870	0.0070		0.1740	0.2030		0.0360		0.0320	0.0540		
12.2										0.0070					
13	0.2160		0.0310	0.0900	0.0010		0.0340	0.1140		0.2990	0.0020	0.0040	0.2330		
13.2	0.0010		0.001							0.0410					
14	0.1650			0.0310	0.0310		0.0060	0.0180		0.2810	0.2480		0.2060		
14.2			0.001							0.0890					
15	0.1630			0.0010	0.3690		0.0020	0.0010		0.0520	0.0250		0.1560		
15.2			0.001	0.001						0.1480					
16	0.0700				0.3160				0.0060	0.0060	0.1650		0.0960		
16.2										0.0280					
17	0.0120				0.2210				0.1030		0.2820		0.0710		
17.2										0.0040					
18	0.0020				0.0500				0.1490		0.1790		0.0460		
18.2										0.0020					
19					0.0030				0.1570		0.0800		0.0580		

Allele	D6S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
20					0.0020				0.1150		0.0160		0.0240		
20.2															
21									0.0220		0.0030		0.0210		
21.2															
22									0.0470				0.0170		
22.2															
23									0.1890				0.0040		
23.2															
24									0.1160						
24.2													0.0050		0.1860
25									0.0710						0.0080
25.2													0.0010		0.0660
26															0.0010
27		0.0010							0.0180						0.0210
28		0.0370							0.0060						0.0090
28.2		0.0060													0.0020
29		0.2070													
29.2		0.0020													
30		0.3740													
30.2		0.0050													
30.3		0.0010													
31		0.1130													
31.2		0.0550													
32		0.0270													
32.2		0.1150													
32.3		0.0010													
33		0.0080													
33.2		0.0430													
34		0.0020													
34.2		0.0020													

Allele	D6S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D6S818	FGA
35		0.0010													
PD	0.9553	0.9248	0.9121	0.8836	0.8692	0.8453	0.9317	0.9257	0.9683	0.9271	0.9260	0.7950	0.9613	0.9149	0.9587
PI	0.0447	0.0752	0.0879	0.1164	0.1308	0.1547	0.0683	0.0743	0.0317	0.0729	0.0740	0.2050	0.0387	0.0851	0.0413
PIC	0.8295	0.7593	0.7316	0.6852	0.6594	0.6264	0.7726	0.7605	0.8585	0.7681	0.7612	0.5629	0.8393	0.7434	0.8355
PE	0.7186	0.5843	0.4861	0.4696	0.4039	0.4069	0.6215	0.5697	0.7226	0.5806	0.5481	0.3232	0.6870	0.5661	0.7147
TPI	3.6232	2.4038	1.8939	1.8248	1.5823	1.5924	2.6596	2.3148	3.6765	2.3810	2.1930	1.3369	3.2468	2.2936	3.5714
Hob	0.8620	0.7920	0.7360	0.7260	0.6840	0.6860	0.8120	0.7840	0.8640	0.7900	0.7720	0.6260	0.8460	0.7820	0.8600
Hex	0.8490	0.7850	0.7690	0.7304	0.7123	0.6719	0.8023	0.7936	0.8729	0.7960	0.7932	0.6245	0.8557	0.7775	0.8531
*P	0.2477	0.1604	0.4778	0.3604	0.7927	0.5794	0.6826	0.9603	0.1179	0.8407	0.4166	0.6701	0.2042	0.8091	0.3595
**P	0.0746	0.1173	0.1765	0.5658	0.5895	0.8080	0.9200	0.9705	0.0973	0.8721	0.3237	0.8162	0.00687	0.8389	0.5148

PD, power of discrimination; PI, power of identity; PIC, polymorphic information content; PE, power of exclusion; TPI, typical paternity index; Hob, observed heterozygosity; Hex, expected heterozygosity; *P, Hard-Weinberg equilibrium, χ^2 test; **P, Hardy-Weinberg equilibrium, exact test based on 10,000 shufflings

표 12. 등록메뉴의 하위메뉴 기능

하위 메뉴	기능	비고
의뢰데이터 개별입력	신원정보를 한 건씩 수동으로 입력	
의뢰데이터 일괄입력	두 건 이상의 정황정보 데이터를 일괄입력	CSV 형식
DNA 일괄입력	두 건 이상의 DNA 프로필정보를 일괄입력	CSV 형식
원본데이터 일괄입력	분석기기에서 나온 Raw-data를 일괄입력	fsa 형식

표 13. 가족코드 번호

가족군			코드식별
가족관계	성별[실종자]		
부 모	부	남	1A
		여	1B
	모	남	1C
		여	1D
자 녀	아들	남	1E
		여	1F
	딸	남	1G
		여	1H

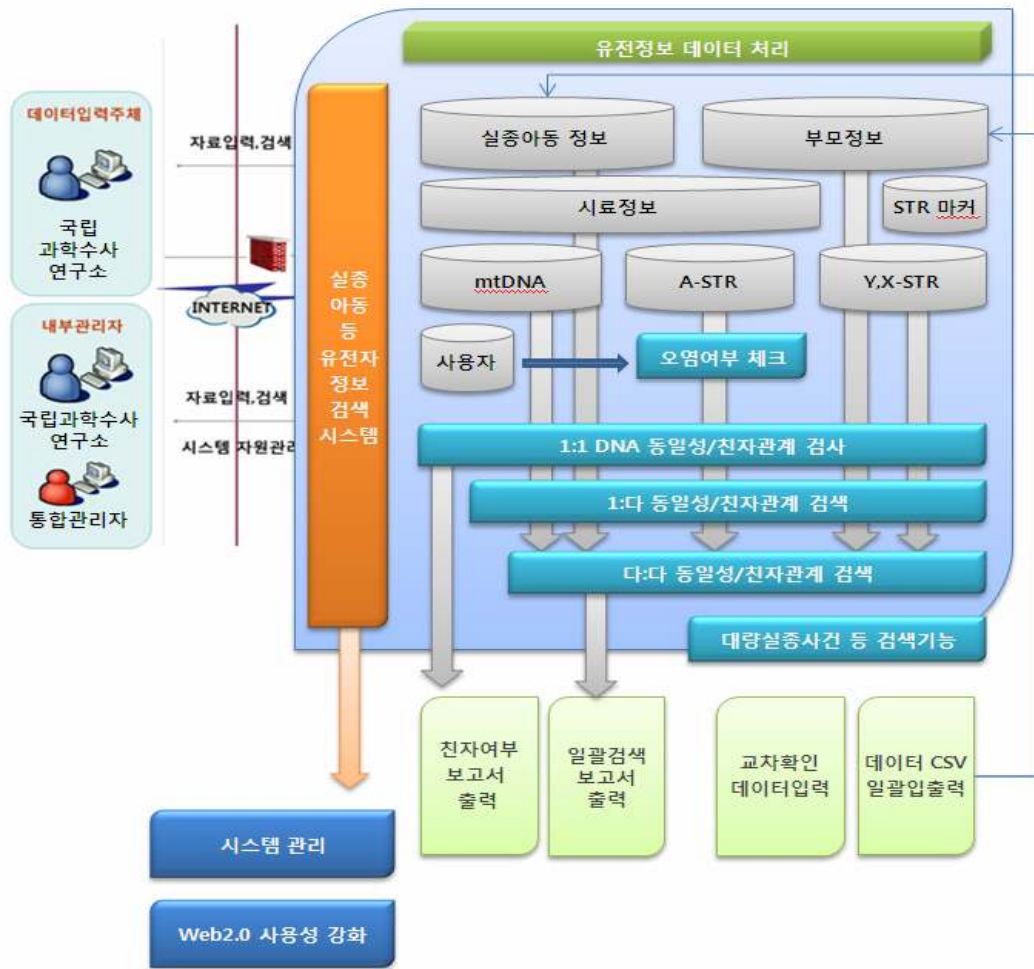


그림 1. 친자확인 DNA 정보 검색시스템 전체 개요



그림 2. 친자확인 DNA 정보 검색시스템 주 화면

친자확인 대상 아동등록

[뒤로가기](#)
[실종아동](#)
[보호자](#)
[변사자](#)
[변사자 보호자](#)

아래 폼을 채우고 제출 버튼을 클릭하세요. * 표시는 필수항목입니다.

* 검체번호:	<input type="text"/>
* 성별:	<input type="text" value="-----"/> ▼
의뢰번호:	<input type="text"/>
요약설명:	<div style="border: 1px dashed gray; height: 100px;"></div>
성년:	<input type="text"/>

[제출](#)
[처음데이터로 돌아가기](#)

그림 3. 친자확인 등록

보호자등록

뒤로가기 실종아동 보호자 변호사 변호사 보호자

아래 폼을 채우고 제출 버튼을 클릭하세요. * 표시는 필수항목입니다.

* 검색번호:	<input type="text"/>
* 성별:	<input type="text" value="-----"/>
의뢰번호:	<input type="text"/>
요약설명:	<div style="border: 1px dashed gray; height: 100px;"></div>
생년:	<input type="text"/>
아동 생년:	<input type="text"/>
실종시기:	<input type="text"/>

제출 처음으로 돌아가기

그림 4. 보호자 등록

A-STR

010
JURE

D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539
D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA	Penta_E
Penta_D	F13A01	F13B	FESFPS	LPL	SE33	XY	
Comment							
최초입력							

저장

처음데이터로 돌아가기

그림 5. A-STR DNA Data 등록화면

mtDNA 010 010

HV1:	과변이 제1영역 (16021-16380)
HV2:	과변이 제2영역 (31-408)
HV3:	과변이 제3영역 (438-574)
Comment:	최초입력

저장
처음데이터로 돌아가기

그림 6. mtDNA 등록화면

일치여부: 예 식별지수: 9.49714725209e+20 불일치수: 0 허용오차: 0					
STR 마커	C-04007992-F		C-07001407-F		일치여부
	Alele 1	Alele 2	Alele 1	Alele 2	
AMEL	X	X	X	X	-
TH01	7	9	7	9	일치
TPOX	8	9	8	9	일치
CSF1PO	11	11	11	11	일치
D3S1358	16	17	16	17	일치
vWA	17	18	17	18	일치
FGA	20	24	20	24	일치
D5S818	7	12	7	12	일치
D13S317	8	9	8	9	일치
D7S820	10	10	10	10	일치
D16S539	9	9	9	9	일치
D8S1179	10	14	10	14	일치
D21S11	30	30	30	30	일치
D18S51	13	14	13	14	일치
D2S1338	19	20	19	20	일치
D19S433	14	14	14	14	일치
Penta_E	12	14	12	14	일치
Penta_D	9	13	9	13	일치

그림 7. 동일성 검사 결과 화면

종합 부모자식관계이부: 예 (아래한글 검증서)

A-STR Result

부모자식관계이부: 예 친부지수: 9,931,036.38813 친부확률: 99.9999899306 % 불일치수: 1 허용오차: 1.0

STR 이커	F10000047-1E-F		C10001217-M		공유여부
	Alele 1	Alele 2	Alele 1	Alele 2	
AMEL	X	X	Y	X	*
TH01	7	9	9	9	배제안됨
TPOX	8	11	8	11	배제안됨
CSF1PO	9	9	9	11	배제안됨
D3S1358	15	15	14	15	배제안됨
vWA	17	19	14	19	배제안됨
FGA	23	23	21	23	배제안됨
D5S818	10	10	9	10	배제안됨
D13S317	8	9	8	8	배제안됨
D7S820	11	12	11	12	배제안됨
D16S539	9	10	10	12	배제안됨
D8S1179	12	14	12	16	배제안됨
D21S11	29	30	30	30	배제안됨
D18S51	11	16	16	16	배제안됨
D251338	20	20	20	22	배제안됨
D19S433	13	14	13	13	배제안됨
Penta_E	11	23	23	23	배제안됨
Penta_D	11	13	11	12	배제안됨

그림 8. 부모자식관계 검사 결과 화면

동일성검색

뒤로가기

검색번호	U04-01-73485-01-F		
검색대상 그룹구분	<input type="checkbox"/> 실종아동 보호자군 <input type="checkbox"/> 변사자 보호자군 <input type="checkbox"/> 실종아동군 <input type="checkbox"/> 변사자군 <input type="checkbox"/> ND 군 <input type="checkbox"/> 대량사망사고군 <input type="checkbox"/> 대량사망사고 보호자군 <input type="checkbox"/> 주요4그룹		
검색대상 신원정보	-성별- 생년 (음선) ~		
A-STR	<input checked="" type="checkbox"/> A-STR search	불일치수 1	허용오차 1 <input type="checkbox"/> 부분검색허용
Y-STR	<input type="checkbox"/> Y-STR search	불일치수 0	허용오차 0 <input type="checkbox"/> 부분검색허용
X-STR	<input type="checkbox"/> X-STR search	불일치수 0	허용오차 0 <input type="checkbox"/> 부분검색허용
mtDNA	<input type="checkbox"/> mtDNA search	<input type="checkbox"/> C-Stretch 영역제외	<input type="checkbox"/> 부분검색허용

검색

처음데이터로 돌아가기

레코드수: 2

검색번호 ↓↑	검사결과	일치율 ↓↑	그룹구분 ↓↑	성별 ↓↑	생년	마등 생년	실종시기	A-STR	Y-STR	X-STR	mtDNA	등록일 ↓↑	등록자 ↓↑
U-08M33310-01-F	일치	16 / 16 (100%)	변사자군	여성	-	-	-	▼				2011-04-26	system
YYC-V-02	일치	16 / 16 (100%)	변사자군	여성	-	-	-	▼				2011-04-26	system

1

◀ 1 페이지 ▶

그림 9. 동일성 검색

● 부모자식관계검색

뒤로가기

검색번호	0601-84800-19		
검색대상 그룹구분	<input type="checkbox"/> 실종아동 보호자군 <input type="checkbox"/> 변사자 보호자군 <input type="checkbox"/> 실종아동군 <input type="checkbox"/> 변사자군 <input type="checkbox"/> ND 군 <input type="checkbox"/> 대량사망사고군 <input type="checkbox"/> 대량사망사고 보호자군 <input type="checkbox"/> 주요4그룹		
검색대상 신원정보	--성별-- 생년 (옵션) ~		
A-STR	<input checked="" type="checkbox"/> A-STR search	불일치수 0	허용오차 0 <input type="checkbox"/> 부분검색허용
Y-STR	<input type="checkbox"/> Y-STR search	불일치수 0	허용오차 0 <input type="checkbox"/> 부분검색허용
X-STR	<input type="checkbox"/> X-STR search	불일치수 0	허용오차 0 <input type="checkbox"/> 부분검색허용
mtDNA	<input type="checkbox"/> mtDNA search	<input type="checkbox"/> C-Stretch 영역제외	<input type="checkbox"/> 부분검색허용

검색 처음데이터로 돌아가기

레코드수: 4

검색번호	검사결과	공유율	전부지수	그룹구분	성별	생년	사망년	실종시기	A-STR	Y-STR	X-STR	mtDNA	등록일	등록자
C-04008333-M	일치	34 / 34 (100%)	10,789,081,048,453,560,964	실종아동군	남성	1960	-	-	✓				2011-04-26	system
R-1001-12094-2-1E-M	일치	34 / 34 (100%)	353,473	변사자 보호자군	남성	1995			✓				2011-07-29	system
C-06014136-M	일치	34 / 34 (100%)	22,058	실종아동군	남성	19	-	-	✓				2011-04-26	system
C-06012519-F	일치	34 / 34 (100%)	2,888	실종아동군	남성	1979	-	-	✓				2011-04-26	system

1

1 페이지

그림 10. 부모자식관계 검색

참고 사이트

Association of Missing & Exploited Children's Organizations (AMECO),

<http://www.amecoinc.org>

BONAPARTE. DVI, <http://Bonaparte DVI www.bonaparte-dvi.com>

ChildSeek Network, <http://www.childseeknetwork.com>

Child Find Canada, <http://www.missingfind.ca>

National Center for Missing & Exploited Children (NCMEC),

<http://www.missingkids.com>

National Crime Information Center (NCIC), <http://www.fbi.gov>

Missing Children Awareness Foundation, <http://www.mcafinc.org>

Missing Children Society of Canada (MCSC), <http://www.mcsc.ca>

Office Of Juvenile Justice Delinquency Prevention (OJJDP),

<http://www.ojjdp.ncjrs.org/index.html>

Vanished Children' s Alliance (VCA), <http://www.vca.org/index.html>

경찰청 실종아동찾기센터, <http://www.182.go.kr>

법무부, <http://www.moj.go.kr>

법제처, <http://www.moleg.go.kr>

실종아동전문기관, <http://www.missingchild.or.kr>

참 고 문 헌

Anderson S, Bankier AT, Barrell BG, Brujin MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schrier, PH Smith AJH, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. Nature. 1981;290:457-465.

Bruce B, Frederick R. B, Arthur J. Eisenberg. Forensic aspects of mass disasters: Strategic considerations for DNA-based human identification. Legal Medicine. 2005;7:230-243.

Brenner CH. Summary of polymorphic STR allele frequencies and Y chromosome haplotype frequencies. Progress in Forensic Genetics. 2000;109-125.

Cho NS, Ahn HJ, Hong SB, Hwang JH, Kim KS, Lim Sk, Kim SH, Lee DS, Kim YS, Kim JH, Shin SC, Lee YJ, Kim NY, Chon BY, Kang PW, Lee KY, Choi DH, Lee YH, Lee YA, Park KW, Kim JJ, Seo JS, Lee YB, Song SB, Kim Y, Han MS. Korean Population Data on 15 STR(Identifiler) loci for Forensic DNA

Profiling. Korean Journal of Forensic Science. 2005;6:142-146

Decorte R, Muslumanoglu MH, Mahieu F, Gilissen A, Cilinger O, Atman C, Artan S, Xiao F, Basaran N, Cassiman J. STR(autosomal and Y chromosome) analysis reveals geographic differences in the turkish population. Progress in forensic Genetics. 2000;8:215-217.

Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. The American Journal of Human Genetics . 1991;49:745-756.

Forster P, Kayser M, Meyer E, Roewer L, Pfeiffer H, Benkmann H, Brinkmann B. Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. Molecular biology and evolution. 1988;15:1108-1114.

Greenberg BD, Newbold JE, Sugino A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene. 1983;21:33-49.

Holland M. M., Parson T.J. Mitochondrial DNA sequence analysis. Validation and use for forensic casework. *Forensic Sci. Rev.* 1999; 11:22-49.

Horst B, Eigel A, Sanguanserm Sri T, Brinkmann B. Human Y-chromosomal STR types in north Thailand. *International Journal of Legal Medicine.* 1999; 112:211-212.

Jobling MA, Pandya A, Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. *International Journal of Legal Medicine.* 1997;110:118-124.

John M. Butler, *Forensic DNA Typing*, Elsevier Academic Press. USA. 2005;Chapter 1,5,9.

Lorente JA, Entrala C, J. Lorente CAM, Lorente AM, Arce B, Carrasco BHF, Budowle B, Villanueva E. *International Journal of Legal Medicine.* 2002;116:187-190

Valdes AM, Slatkin M, Freimer NB. Allele frequencies at microsatellite

loci: The stepwise model revisited. Genetics 1993;133:737-749

van Dongen CJ, Slooten K, Slagter M, Burgers W, Wiegerinck W. Bonaparte: Application of new software for missing persons program. Forensic Science International: Genetics supplement series 3. 2011; 119-120

김종우, 한국의 실종아동 찾기 사업의 발전방안 연구. 2006

김성천, 실종아동사업 발전방안에 관한 연구. 2008

Abstract

Development of DNA Database searching and paternity analysis system for missing children

Young Eun Kim

Department of Biology,

Sungshin Women's University, Seoul.

DNA markers that can recognize individuals within human genome, such as autosomal STR, DNA markers within human genome, such as autosomal STR, sex chromosomal STR, and mitochondrial DNA sequences can identify individuals and can be used to test paternity. As the DNA technology advances, the amount of STR data has increased. However, because of the lack of systematic database system, DNA data has been managed with Access Program and the situational information has been managed with Excel Program. It was a dual structure. With the enactment of “The Law Regarding the Protection and Support of Missing Children, Etc.” in 2005, the DNA testing institutions were able to establish a database for the

DNA testing results. Therefore, this study established “the DNA Information Search System for Paternity Tests.”

The paternity test DNA information system consists of data of 4 groups. They were missing children group, parents group, group of dead bodies, and group of their protectors. This system has an overall search function that can compare the children with parent groups one by one, and with all profiles within the database. It can also accomplish the comparative search between groups. The search results are shown with the standard error and paternity index.

After the search the candidates were concluded according to the standard deviation considering the mutations of comparative DNA. It minimized the time and efforts to deduce the familial relationships. This system will increase the efficiency of paternity test search tasks by considering the convenience of the users. In addition, the DNA analysis results will contribute to finding the paternity quickly, and to responding accurately to the social issues by reflecting the recent scientific statistical calculation methods and providing significant information.