

유전자 분석 자료에 의한 친자 및 혈연관계 분석시스템 개발 및 활용

구교찬, 김선욱*
¹단국대학교 산업공학과

Development and Applications of A Paternity and Kinship Analysis System Based on DNA Data

Kyo-Chan Koo¹, Sun-Uk Kim^{1*}

¹Department of Industrial Engineering, Dankook University

요약 최근 실종자, 변사자, 미아 등의 유전자 분석 자료는 지속적으로 증가하고 있으나, 현재 친자확인을 위한 통계학적 계산은 대부분 수기에 의하거나 엑셀을 통해서 이루어지고 있다. 따라서 유전자 분석 자료 중 상염색체 Short Tandem Repeat (STR)을 체계적으로 관리하고 효과적으로 분석할 수 있는 소프트웨어의 개발이 필요하다. 친자관계 및 혈연관계를 다양한 옵션 하에서 용이하게 분석하는 웹 기반 유전자자료 분석시스템이 광범위한 테스트 없이 약 20개월의 연구를 통해서 개발되었다. 친자관계 분석을 위해서 부계지수 계산 알고리즘을 사용하였고, 혈연관계 분석을 위해서 Identity by descent (IBD) 공식을 사용하였다. 이 시스템은 실제 데이터를 기반으로 혈연관계지수와 친자확률이 검증됨으로써 신뢰성이 확보됨은 물론, 대량 재난재해 시 발생될 유전자 분석 자료의 관리 및 분석에 효과적으로 이용될 수 있을 것이다. 이 외에도 본 시스템은 데이터베이스와 알고리즘의 통합 환경, 사용자 중심 인터페이스, 프로세스 자동화 등 고급기능을 포함한다.

Abstract Recently, DNA data of missing person, killed person, and missing child continue to increase but most of statistical calculation for paternity confirmation is being done through manual methods or Excel. Therefore, we need development of a software which is able to facilitate both systematic management and effective analysis of Short Tandem Repeat (STR) derived from DNA data. Without extensive testing, through a twenty-month study was developed a web-based system which performs paternity analysis and kinship analysis easily based on the various options. The former uses an existing algorithm for paternity index and the latter does Identity by descent (IBD) formula. Due to our system validated over real datasets in terms of likelihood ratio and probability of paternity, it ensures increased reliability as well as effective management and analysis of DNA data in mass disaster. In addition, it includes advanced features such as an integrated environment, user-centered interface, process automation and so on.

Keywords : DNA, Kinship analysis, Paternity analysis, Short Tandem Repeat (STR), Web-based system

1. 서론

최근 실종자, 변사자, 미아 등의 유전자 분석 자료는 지속적으로 증가하고 있지만, 부계지수 등 친자 및 혈연

관계 확인을 위한 통계학적 계산은 대부분 수기 또는 엑셀을 통해서 이루어지고 있어, 입력시간이 오래 걸릴 뿐 아니라 입력과정에서 오류가 발생할 수 있다. 또한 사용자가 마커(Marker), 분석키트(Analysis Kit), 빈도표

*Corresponding Author : Sun-Uk Kim(Dankook University)

Tel: +82-41-550-3570 email: kimsunuk@dankook.ac.kr

Received July 2, 2015

Accepted October 8, 2015

Revised (1st August 5, 2015, 2nd August 12, 2015, 3rd August 17, 2015, 4th August 31, 2015)

Published October 31, 2015

(Frequency Table)를 관리를 할 수 없고, 사전확률 (Pre-Probability), 부분매칭(Partial Matching), 돌연변이 (Mutation) 등의 다양한 분석옵션을 적용할 수 없어 분석이 제한적이다.

현재 경찰청에서 개발한 미아찾기를 위한 시스템이 있지만, 사용이 불편하고 시스템의 제한 때문에 유지관리가 어렵다. 최근에 국방부 조사본부에서는 6.25전사자 유해에서 조사된 DNA 정보를 기반으로 하여 혈연관계를 확인하는 소프트웨어를 개발하여 발표하였다[1]. 해외에서 친자 및 혈연관계 확인을 위한 소프트웨어가 시판중이지만, 비용이 비싸고 국내에서 사용되고 있는 시스템과 호환시키는 것도 쉽지 않다[Table 1][2]. 따라서 지속적으로 증가하는 유전자 분석 자료를 체계적으로 관리하고 효과적으로 검증할 수 있는 시스템 개발이 필요하다[3,4].

이에, 본 연구에서는 친자 및 혈연관계 분석에 필요한 상염색체 STR 자료와 다양한 옵션들을 데이터베이스화 하고, 부계지수 계산 알고리즘과 IBD 공식을 사용하여 통계학적 검증이 가능하도록 통합할 뿐만 아니라 웹서비스가 가능한 사용자 중심 친자 및 혈연관계 분석시스템을 개발하였다.

Table 1. Tabular evaluation of program function.

| Software name | Update | Last Trio | SS Mother/related | Complicated | Kinship | Mutation | allelic N u l l | Fraternity | U s e r |
|---|--------|-----------|-------------------|-------------|---------|----------|-----------------|------------|---------|
| DNAStat | 2007 | ● | ● | | | ● | ● | ● | |
| DNA-View | 2008 | ● | ● | ● | ● | ● | ● | ● | |
| EasyDNA | 2008 | ● | ● | ● | ● | | | | ● |
| EasyPat | 2000 | ● | ● | | ● | | | | |
| familias | 2006 | ● | ● | ● | ● | ● | ● | ● | |
| GenoProof | 2008 | ● | ● | | ● | ● | | | ● |
| Genotype | 2008 | ● | ● | ● | ● | ● | ● | ● | |
| Hugin | 2004 | ● | ● | ● | ● | ● | ● | ● | |
| PatCan | 2003 | ● | ● | | ● | | | | ● |
| Patern | 1996 | ● | | | | | | | |
| Paternity Index | 2007 | ● | ● | ● | ● | ● | ● | ● | ● |
| PatPCR | 2003 | ● | ● | | | | | | ● |
| [Calculation of the pedigree probability] | 2004 | ● | ● | ● | ● | | | | ● |

2. 연구방법

2.1 친자 및 혈연관계 확인을 위한 표준 STR 마커

STR은 돌연변이율이 높고 multiplex 분석이 용이하기 때문에 유전자 감식 분야에서 매우 널리 이용되고 있다[5]. FBI에서 1997년에 U.S. national DNA database (NDNAD)와 국제 범죄자 데이터베이스 구축을 위해 지정한 표준 마커인 13 CODIS (Combined DNA Index System) 마커를 포함하여[6,8], 국제적으로 상용화되어 가장 널리 쓰이고 있는 AmpFISTR Identifiler (Applied Biosystems, USA)와 PowerPlex 16 (Promega, USA) 등에 포함된 17개 마커를 표준 STR 마커로 선정하였다 [Table.2].

Table 2. STR markers of 13 CODIS, Identifiler and PowerPlex 16

| 13 CODIS | Identifiler | PowerPlex 16 |
|----------|-------------|--------------|
| 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 |
| - | D2S1338 | Penta D |
| - | D19S433 | Penta E |
| - | Amelogenin | Amelogenin |

Identifiler를 기준으로 하고, PowerPlex 16의 Penta D 와 Penta E 마커를 부가적으로 포함하였다. 또한, 새로운 키트 시스템의 추가 및 관리를 가능하게 함으로써 향후 식별력을 높이기 위해 새로운 마커를 계산에 이용할 수 있도록 하였다.

2.2 지수계산 알고리즘

2.2.1 친자관계 지수계산

친자관계 확인 결과의 통계학적인 검증을 위해 한쪽 부모와 자녀(Duo), 양쪽 부모와 자녀(Trio)사이에는 부계지수(Paternity Index, PI)를 계산하는 알고리즘이 활용되었다[Table 3,4][9,10].

Table 3. Combinations of paternal child genotypes (Duo)

| Genotype Child | Genotype Alleged Father | PI |
|----------------|-------------------------|------------------------------|
| AA | AA | $\frac{1}{P_A}$ |
| AA | AB | $\frac{1}{2P_A}$ |
| AB | AA | $\frac{1}{2P_A}$ |
| AB | AB | $\frac{P_A + P_B}{4P_A P_B}$ |
| AB | AC | $\frac{1}{4P_A}$ |

Table 4. Combinations of maternal, paternal and child genotypes (Trio)

| Genotype Mother | Genotype Child | Genotype Alleged Father | PI |
|-----------------|----------------|-------------------------|--------------------------|
| AA | AA | AA | $\frac{1}{P_A}$ |
| AB | | | |
| BB | | | |
| BC | AB | AB | $\frac{1}{2P_A}$ |
| AA | | | |
| AB | | | |
| AC | AA | AB | $\frac{1}{2P_A}$ |
| BB | | | |
| BC | | | |
| BC | AB | AB | $\frac{1}{2P_A}$ |
| CC | | | |
| CD | | | |
| AB | AB | AA | $\frac{1}{P_A + P_B}$ |
| | | AB | $\frac{1}{2(P_A + P_B)}$ |
| | | AC | $\frac{1}{2(P_A + P_B)}$ |

예를 들어 STR마커 중 D3S1358의 두 대립유전자 (Allele)가 자(15,16) / 부(15,16) 이라면 A=15, B=16이다. Jin 등[11]의 연구에 따르면 각 대립유전자의 빈도 $P_{15} = 0.4$, $P_{16} = 0.321$ 임을 알 수 있다. Table 2에서 자녀 유전자형이 AB이고 추정부 유전자형이 AB이므로, 부계 지수의 공식인 $(P_A + P_B)/(4P_A P_B)$ 를 적용하면 $(0.4+0.321)/(4 \times 0.4 \times 0.321)=1.403$ 임을 구할 수 있다. 각 마커의 PI값을 계산한 후 각 값을 서로 곱한 결과를 결합친부지수(Combined Paternity Index, CPI)라 한다. 결합친부지수로 친부확률(Probability of Paternity, PP)을 구하는 공식은 다음과 같다.

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

2.2.2 혈연관계 지수계산

혈연관계 확인 결과의 통계학적인 검증을 위한 지수

계산에는 IBD에 의한 공식을 사용하였다[Table 5,6][10,12].

Table 5. Identity by descent coefficients $\{\Phi_2, \Phi_1, \Phi_0\}$

| Relation categories | 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 | $[P_2(xy) \times \Phi_2] + [P_1(xy) \times \Phi_1] + [P_0(xy) \times \Phi_0]$ | | |

Table 6. Kinship Formulas

| allele 1 | allele 2 | Frequency |
|----------|----------|---|
| AB | AB | $\Phi_2 + 0.5\Phi_1(P_A + P_B) + 2\Phi_0 P_A P_B$ |
| 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 | $\Phi_0 P_A^2$ |

혈연관계 지수계산 방법은 친자관계 지수계산방법과 동일하게 두 대립유전자를 비교한 후 빈도를 구하고 Table 5,6을 참고하여 혈연관계지수를 구할 수 있다[13]. 단, 친자관계 확인과 다른 점은 각 혈연관계에 따라 공식과 Φ 를 적용하여 각 마커별로 값을 구한다는 것이다.

예를 들어 STR마커 중 D5S818 두 대립 유전자가 (11,11) / (11,12) 이라면 A=11, B=12이고 대립유전자 빈도 $P_{11} = 0.312$, $P_{12} = 0.242$ 이다. 이때 Full sibs의 경우 $\Phi_1 P_B + 2\Phi_0 P_A P_B$ 의 공식을 적용하면 $0.5 \times 0.242 + 2 \times 0.25 \times 0.312 \times 0.242 = 0.159$ 이며, Unrelated는 $2 \times 1 \times 0.312 \times 0.242 = 0.151$ 로 혈연관계지수(likelihood ratio)를 계산하면 $0.159/0.151=1.053$ 이다. 각 마커별로 구해진 값들을 모두 곱한 후에 Unrelated의 곱해진 값으로 나눈 값이 혈연관계지수이며, 각각의 혈연관계확률은 {혈연관계지수/(혈연관계지수+1)} $\times 100$ 이며, 각 혈연관계별로 계산된 확률 중 가장 높은 수치의 결과를 확인하여 혈연관계를 결정한다.

2.3 시스템 개발도구

본 연구에서는 ‘MySQL’ 데이터베이스와 ‘Java’, ‘Jsp’ 프로그래밍언어가 개발도구로 사용되었다. 전자는

상염색체 STR 유전자 분석 자료, 마커, 분석키트, 빈도표를 관리하는데 이용되며 지수계산을 위한 기본정보로 활용된다. 후자는 지수계산 알고리즘을 구현하고 데이터베이스와 연동하여 시스템을 통합하기 위한 프로그래밍 언어로 이용된다.

본 시스템은 ‘Apatch Tomcat’ 서버에서 구동되며, MS Windows XP 이상의 운영체제에서 작동되도록 하였다.

3. 연구결과

3.1 시스템 구현

본 시스템은 2010년 4월부터 2011년 12월까지 정부의 K기관 지원으로 프로토타입이 개발되었으며, 최근까지 수정 및 보완되어 지금은 안정적으로 운영되고 있다.

시스템은 Home, Management, Analysis, Result, Administration 이상 5개의 주메뉴와 My Page, Help 이상 2개의 부메뉴 구성되어 있다[Fig 1].

Fig 1. Main menu of Analysis system



Management 메뉴에는 Sample, Marker, Analysis Kit, Frequency Table 이상 4개의 서브메뉴로 구성되어 있으며 초기에는 샘플번호별 등록정보를 확인할 수 있으며, 샘플을 등록, 수정, 삭제할 수 있다. Marker, Analysis Kit, Frequency Table 메뉴는 시스템에 등록되어 있는 Marker, Analysis Kit, Frequency Table을 확인할 수 있다. Analysis 메뉴에는 Paternity Analysis-Duo, Paternity Analysis-Trio, Kinship Analysis 이상 3개의 서브메뉴로 구성되어 있다. 서브메뉴에 분석옵션을 두어 결과 값에 제한을 두었다. 옵션에는 Pre-Probability, Cut off PP, Cut off CPI, Partial Matching (Mutation

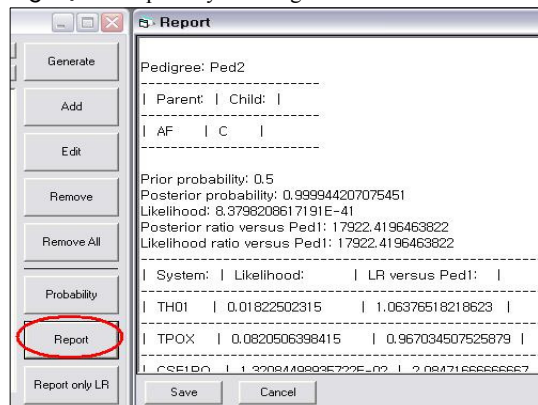
Num/Allele Range), Analysis Kit, Frequency Table이 있으며, 각 옵션을 선택, 입력, 체크하여 사용한다. Result 메뉴의 서브메뉴는 Analysis 메뉴와 동일하며 분석결과를 보여준다. Administration 메뉴에서는 시스템 사용자, 마커, 분석키트, 빈도표 등을 등록, 수정, 삭제할 수 있다.

3.2 시스템 검증

지금까지 알려진 유전자분석 소프트웨어는 상용화된 것들과 무료로 공개된 소프트웨어 등 여러 가지가 있고 그 중 대표적인 것이 DNA-View 라는 소프트웨어가 있다. Drábek은 DNA-View를 포함한 대표적인 13개의 유전자분석 소프트웨어에 대해 validation을 실시하였으며 likelihood ratio 및 kinship analysis 등을 비교하였다. 그 결과 DNA-View기준으로 familias와 Paternity Index 소프트웨어가 가장 적합하다고 판정하였다[2,14]. 그 중 본 시스템에 대한 검증은 비교적 간단하고 연구대상과 환경이 유사한 familias와의 비교를 통해 실시되었다.

Commercial off-the-shelf (COTS) software인 familias는 calibration과 테스트 (EN/ISO/IEC 17025:2005)를 거친 software로 likelihood ratio와 친부지수 분석에 널리 사용되고 있다. 본 연구의 검증을 진행하기 위해 사용된 빈도표는 Jin 등[11]의 연구에서 사용된 자료를 입력하였다. familias에는 데이터베이스를 저장하는 기능이 없기 때문에 특정 대상을 직접 비교하는 방법으로 실행했다. 혈연관계분석결과 주요 지표값인 likelihood ratio와 확률은 본 시스템[Fig 4]과 familias시스템[Fig 2]이 서로 일치하는 것으로 나타났다.

Fig 2. Kinship analysis using familias



3.3 시스템 활용사례

본 시스템에서는 GeneMapper로부터 출력된 유전자 분석 결과의 입력을 자동화하여 입력과정에서의 오류를 방지하였고, 사용자의 편의성을 증가시켰다. 웹 기반의 친숙한 인터페이스로 사용자가 쉽게 익히고 사용이 가능하도록 구성하였다. 입력된 샘플 간에 일대일, 부분 샘플을 선택, 전체 샘플 등에 대해 분석이 가능하도록 하였다. 또한 사용자가 마커를 추가하고 이를 이용하여 새로운 분석키트로 구성할 수 있게 함으로써 분석에 확장성도 높였다. 빈도표를 사용자가 구성하고 여러 집단 또는 여러 종류의 빈도표를 입력하고 분석할 때 선택할 수 있다. 친자 및 혈연관계 분석 시 돌연변이의 범위와 개수를 설정하고 돌연변이가 있을 것으로 판단될 경우 Brenner's method[15]를 적용하여 계산되도록 하였다.

다음은 이러한 사용자 인터페이스와 프로세스 자동화에 기초하여 주요기능인 친자관계 및 혈연관계분석을 예시한다.

친자관계분석은 Analysis 하위메뉴인 Paternity Analysis-Duo를 이용하여 분석하며, 분석옵션에는 Pre-Probability = 10, Cut off CPI = 100, Analysis Kit = PowerPlex16, Frequency Table = Korean165[11]을 사용하여 수행되었다. 여기에서 이용된 옵션은 흔히 사용하는 옵션이며, 옵션은 사용자 편의, 정밀도, 분석결과 등에 따라 변경할 수 있다. Fig 3는 친자관계분석 옵션에 따른 부계지수 및 친자확률이며, 친자확률이 99.5992%이므로 친자관계로 판정할 수 있다.

Fig 3. Result of Paternity Analysis-Duo

| Marker | Parent | Child | a | b | Pa | Pb | Type | Formula | PI |
|--------------------------------|-----------|-----------|-----------|------|-------------------------------|-------|----------|---------|--------|
| D3S1358 | 15 / 21 | 15 / 16 | 15 | 21 | 0.4 | 0.015 | abac | 1/4Pa | 0.6249 |
| TH01 | 7 / 7 | 6 / 7 | 7 | 6 | 0.288 | 0.112 | aaab | 1/2Pa | 1.7361 |
| D21S11 | 29 / 33.2 | 29 / 32.2 | 29 | 33.2 | 0.252 | 0.058 | abac | 1/4Pa | 0.9920 |
| D18551 | 15 / 17 | 14 / 15 | 15 | 17 | 0.173 | 0.076 | abac | 1/4Pa | 1.4450 |
| Penta_E | 14 / 19 | 10 / 14 | 14 | 19 | 0.085 | 0.046 | abac | 1/4Pa | 2.9411 |
| D5S818 | 11 / 13 | 12 / 13 | 13 | 11 | 0.136 | 0.312 | abac | 1/4Pa | 1.8382 |
| D13S317 | 9 / 11 | 8 / 9 | 9 | 11 | 0.124 | 0.221 | abac | 1/4Pa | 2.0161 |
| D7S820 | 8 / 12 | 8 / 11 | 8 | 12 | 0.118 | 0.188 | abac | 1/4Pa | 2.1186 |
| D16S539 | 9 / 13 | 9 / 9 | 9 | 13 | 0.294 | 0.085 | abaa | 1/2Pa | 1.7006 |
| CSF1PO | 10 / 11 | 10 / 12 | 10 | 11 | 0.239 | 0.242 | abac | 1/4Pa | 1.0460 |
| Penta_D | 10 / 13 | 13 / 13 | 13 | 10 | 0.1 | 0.173 | abaa | 1/2Pa | 4.9999 |
| vWA | 14 / 19 | 14 / 14 | 14 | 19 | 0.212 | 0.07 | abaa | 1/2Pa | 2.3587 |
| D8S1179 | 12 / 13 | 10 / 12 | 12 | 13 | 0.124 | 0.224 | abac | 1/4Pa | 2.0161 |
| TPOX | 9 / 11 | 8 / 11 | 11 | 9 | 0.333 | 0.121 | abac | 1/4Pa | 0.7507 |
| FGA | 23 / 23.2 | 23 / 23 | 23 | 23.2 | 0.255 | 0.006 | abaa | 1/2Pa | 1.9607 |
| CPI (Combined Paternity Index) | | | 2236.6719 | | PP (Probability of Paternity) | | 99.5992% | | |

혈연관계분석은 Analysis 하위메뉴인 Kinship Analysis를 이용하여 분석하며, 분석옵션은 친자과계분

석과 동일하게 수행하였다. Fig 4은 혈연관계지수 및 혈연관계확률이며, Full-sib의 확률이 가장 높고, 99.9999%이므로 형제-자매간의관계로 판정할 수 있다.

Fig 4. Result of Kinship Analysis

| Marker | allele 1 | allele 2 | a (Pa) | b (Pb) | c (Pc) | d (Pd) | Type Formula | Relation | | | | |
|---------------------|-----------|-----------|---------------|-----------------|---------------|--------------|---------------------------------|--------------|-----------|--------------|--------------|------------|
| | | | | | | | | PC | FS | HS | FC | UR |
| D3S1358 | 15 / 15 | 16 / 17 | 15 (0.4) | 16 (0.321) | 17 (0.182) | 0 (0.015) | abac 2Φ0PbPc | 0.000 | 0.029 | 0.058 | 0.088 | 0.117 |
| TH01 | 7 / 8 | 7 / 8 | 7 (0.288) | 8 (0.047) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.167 | 0.341 | 0.097 | 0.062 | 0.027 |
| D21S11 | 29 / 32.2 | 29 / 32.2 | 29 (0.252) | 32.2 (0.088) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.170 | 0.346 | 0.107 | 0.076 | 0.044 |
| D18551 | 13 / 15 | 13 / 15 | 13 (0.212) | 15 (0.173) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.192 | 0.365 | 0.133 | 0.103 | 0.073 |
| Penta_E | 5 / 18 | 5 / 18 | 5 (0.067) | 18 (0.073) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.070 | 0.287 | 0.040 | 0.025 | 0.010 |
| D5S818 | 12 / 12 | 12 / 12 | 12 (0.242) | 0 (0.015) | 0 (0.015) | 0 (0.015) | aaaa Φ2-Φ1Pa-Φ0Pa^2 | 0.242 | 0.386 | 0.150 | 0.104 | 0.059 |
| D13S317 | 9 / 12 | 9 / 12 | 9 (0.124) | 12 (0.194) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.159 | 0.342 | 0.104 | 0.076 | 0.048 |
| D7S820 | 11 / 11 | 11 / 11 | 11 (0.373) | 0 (0.015) | 0 (0.015) | 0 (0.015) | aaaa Φ2-Φ1Pa-Φ0Pa^2 | 0.373 | 0.471 | 0.156 | 0.198 | 0.139 |
| D16S539 | 9 / 9 | 9 / 9 | 9 (0.294) | 0 (0.015) | 0 (0.015) | 0 (0.015) | aaaa Φ2-Φ1Pa-Φ0Pa^2 | 0.294 | 0.419 | 0.190 | 0.138 | 0.086 |
| CSF1PO | 11 / 12 | 11 / 12 | 11 (0.242) | 12 (0.394) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.318 | 0.457 | 0.254 | 0.223 | 0.191 |
| Penta_D | 12 / 13 | 12 / 13 | 12 (0.146) | 13 (0.1) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.123 | 0.319 | 0.076 | 0.053 | 0.029 |
| vWA | 18 / 19 | 18 / 19 | 18 (0.194) | 19 (0.07) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.132 | 0.323 | 0.080 | 0.053 | 0.027 |
| D8S1179 | 10 / 15 | 10 / 15 | 10 (0.142) | 15 (0.158) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.150 | 0.336 | 0.097 | 0.071 | 0.045 |
| TPOX | 8 / 11 | 8 / 11 | 8 (0.449) | 11 (0.333) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.391 | 0.520 | 0.345 | 0.322 | 0.299 |
| FGA | 18 / 22 | 18 / 22 | 18 (0.015) | 22 (0.197) | 0 (0.015) | 0 (0.015) | abab Φ2+0.5Φ1(Pa+Pb)+2Φ0PaPb | 0.106 | 0.304 | 0.056 | 0.031 | 0.006 |
| Relation | | | | | | | | Parent-Child | Full-sib | Half-sib | First Cousin | Unrelated |
| Multiplication | | | | | | | | 0.0 | 2.3462E-8 | 7.0942E-15 | 1.0148E-16 | 4.0908E-20 |
| Index(VS Unrelated) | | | | | | | | 0.0 | 5.7353E11 | 1.73418.1037 | 2479.6701 | 1.0 |
| Probability | | | | | | | | 0.0% | 99.9999% | 99.9994% | 99.9956% | 50.0% |

4. 결론

본 연구에서는 친자확인 결과의 통계학적 검증을 통해서 과학적인 신뢰성과 정확성을 확보하고, 대량 재난·재해를 대비해서 방대한 유전자 분석 자료들을 대상으로 친자 및 혈연관계의 신속한 확인이 가능한 시스템을 개발하고자 하였다. 따라서 본 연구에서는 친자 및 혈연관계 확인 결과의 통계학적 검증을 위해, Lee 등[10]의 연구에 나오는 방법과 IBD 공식을 이용하여 유전자 분석 자료들 간에 비교 및 계산을 수행하는 지수계산 알고리즘을 이용하고 이를 활용하여 친자 및 혈연관계 분석시스템을 개발하였다. 본 시스템은 개인정보 및 유전자 분석 자료의 입력과정을 간소화하고, 전산망에서 출력된 자료를 직접 입력함으로써 업무 효율성과 정확성을 향상할 수 있도록 하였다. 또한, 사용자 편의를 위한 인터페이스로 업무 효율을 향상시켰으며, 실종자·변사자·미아찾기를 위한 데이터베이스를 쉽고 효율적으로 구축하고 관리할 수 있도록 하였다.

현재 국립과학수사연구원에서는 실종자·변사자·미

아의 DNA 프로필이 모두 별도의 시스템에서 관리되고 있어 활용이 매우 불편하고 비효율적이다. 본 시스템을 활용함으로써 실종자·변사자·미아에 대한 유전자 프로필의 정확하고 편리한 입력 및 관리를 통해서 업무 효율 및 신뢰성을 향상시키고 효율적인 데이터베이스의 관리가 가능해질 수 있다.

본 시스템을 활용함으로써 대량 재난·재해를 대비한 시스템을 구축할 수 있다. 미국에서 9.11 테러와 같은 대형 재난이 발생하였을 때 엄청난 양의 유전자 정보가 발생하였는데, 컴퓨터에 의해 자동화된 처리 방식이 큰 도움이 되었다. 이처럼 대형 재난·재해 시 발생하는 방대한 양의 유전자 정보를 기존 방식으로 처리하기는 쉽지 않다. 따라서 언제 발생할지 모르는 대형 재난·재해를 대비하여 전산화된 자동화 시스템을 구축함으로써 많은 양의 유전자 분석 자료를 쉽게 관리하고, 친자 및 혈연관계를 신속, 정확하게 분석하여 실종자 및 변사자의 가족을 빠르게 찾아줄 수 있을 것이다.

Reference

- [1] S. B. Hong, J. Y. Kim, H. J. Park, H. J. Ahn, "Database searching and kinship analysis system of STR and mtDNA data in Korean war remains", Korean journal of forensic science, Vol. 11, No. 1, pp. 12-18, 2010.
- [2] J. Drábek, "Validation of software for calculating the likelihood ratio for parentage and kinship", Forensic Science International: Genetics, Vol. 3, No. 2, pp. 112-118, 2008.
DOI: <http://dx.doi.org/10.1016/j.fsigen.2008.11.005>
- [3] J. A. Riancho, M. T. Zarrabeitia, "A Windows-based software for common paternity and sibling analyses", Forensic Science International, Vol. 135, No. 3, pp. 232-234, 2003.
DOI: [http://dx.doi.org/10.1016/S0379-0738\(03\)00217-2](http://dx.doi.org/10.1016/S0379-0738(03)00217-2)
- [4] K. F. Goodnight, D. C. Queller, "Computer software for performing likelihood test of pedigree relationship using genetic markers", Molecular Ecology, Vol. 8, No. 7, pp. 1231-1234, 1999.
DOI: <http://dx.doi.org/10.1046/j.1365-294x.1999.00664.x>
- [5] J. M. Butler, "Genetics and genomics of core short tandem repeat loci used in human identity testing", Journal of Forensic Sciences, Vol. 51, No. 2, pp. 253-265, 2006.
DOI: <http://dx.doi.org/10.1111/j.1556-4029.2006.00046.x>
- [6] J. M. Butler, "Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers", Academic Press, 2005.
- [7] P. Gill, "Role of short tandem repeat DNA in forensic casework in the UK-past, present, and future perspectives", Biotechniques, Vol. 32, No. 2, pp. 366-385, 2002.
- [8] M. A. Jobling, P. Gill, "Encoded evidence: DNA in forensic analysis", Nature Reviews Genetics, Vol. 5, No. 10, pp. 739-751, 2004.
DOI: <http://dx.doi.org/10.1038/nrg1455>
- [9] J. S. Buckleton, C. M. Triggs, S. J. Walsh, "Forensic DNA Evidence Interpretation", CRC Press, 2005.
- [10] H. J. Lee, J. W. Lee, G. R. Han, J. J. Hwang, "Motherless case in paternity testing", Forensic science international, Vol. 114, No. 2, pp. 57-65, 2000.
DOI: [http://dx.doi.org/10.1016/S0379-0738\(00\)00293-0](http://dx.doi.org/10.1016/S0379-0738(00)00293-0)
- [11] H. J. Jin, K. D. Kwak, S. B. Hong, Y. H. Cho, M. S. Han, W. Kim, "Forensic Genetic Analysis for the PowerPlex-16 System in the Korean Population", Gene & Genomics, Vol. 29, No. 4, pp. 489-496, 2007.
- [12] M. S. Blouin, "DNA-based methods for pedigree reconstruction and kinship analysis in natural population", Trends in Ecology & Evolution, Vol. 18, No. 10, pp. 503-511, 2003.
DOI: [http://dx.doi.org/10.1016/S0169-5347\(03\)00225-8](http://dx.doi.org/10.1016/S0169-5347(03)00225-8)
- [13] F. Dai, D. E. Weeks, "Ordered Genotypes: An extended ITO Method and a general formula for genetic covariance", The American Journal of Human Genetics, Vol. 78, No. 6, pp. 1035-1045, 2006.
DOI: <http://dx.doi.org/10.1086/504045>
- [14] T. Egeland, P. F. Mostad, B. Mevåg, M. Stenersen, "Beyond traditional paternity and identification cases: Selecting the most probable pedigree", Forensic Science International, Vol. 110, No. 1, pp. 47-59, 2000.
DOI: [http://dx.doi.org/10.1016/S0379-0738\(00\)00147-X](http://dx.doi.org/10.1016/S0379-0738(00)00147-X)
- [15] C. H. Brenner, "Multiple mutations, covert mutations and false exclusions in paternity casework", International Congress Series, Vol. 1261, pp. 112-114, 2004.
DOI: [http://dx.doi.org/10.1016/S0531-5131\(03\)01843-0](http://dx.doi.org/10.1016/S0531-5131(03)01843-0)

구 교 찬(Kyo-Chan Koo)

[정회원]



- 2010년 2월 : 단국대학교 산업공학과(공학석사)
- 2012년 9월 ~ 현재 : 단국대학교 산업공학과 박사과정

<관심분야>

인공지능 및 전문가시스템, Bioinformatics, 인간공학

김 선 욱(Sun-Uk Kim)

[정회원]



- 1981년 2월 : 고려대학교 산업공학과(공학석사)
- 1990년 7월 : Oregon State University 산업 및 제조공학과(공학박사)
- 1991년 3월 ~ 현재 : 단국대학교 산업공학과 교수

<관심분야>

정보시스템, 인공지능 및 전문가시스템, 인간공학