Paternity Test using Genetics

1st Youssif Assem Mondy
Misr International University
youssifassem18015221@miuegypt.edu.eg

2nd Mohamed Moataz Kamel
Misr International University
mohamedmoataz1810432@miuegypt.edu.eg

3rd Kareem Ehab Abdelrazek
Misr International University
kareem1802405@miuegypt.edu.eg

4th Ahmed Gamal
Misr International University
ahmed1804264@miuegypt.edu.eg

4th Mohamed Akram Yehia
Misr International University
Mohamed1700772@miuegypt.edu.eg

5th Dr. Ashraf abdelraouf
Misr International University
ashraf.raouf@miuegypt.edu.eg

6th Eng. Ahmed Hazem
Misr International University
ahmed.abo-elsoud@miuegypt.edu.eg

***Abstract*—The most accurate aspect in determining the parents of a child is their DNA. One of the most significant phenomena that have spread around the world are street children, orphaned children, people who can be criminals, or men who dispute paternity to a child, we want to develop a method that analyzes the whole genome or alleles. The system will utilize numerous algorithms to compare it to all of the individuals in our data set, checking for resemblance with each one, and then telling them what family this child may be related to. In addition, the algorithm can determine whether or not the child is connected to the father. Furthermore, we intend to expand our system in the future to help the police with their criminal investigations and help the court with the proof of paternity.**

I. Introduction

One of the key aspects in studying humans nature are human genes. Every person on the planet has a unique fingerprint. We can prove some Key findings using a person's genes. For instance, criminal cases, paternity cases, street children, orphans, etc. Many fathers refuse to acknowledge the paternity of their children. In these circumstances, the court orders a laboratory test to determine whether or not the father is the biological father of the child. According to the study entitled "Forensic DNA Profiling of a Population Sample from Upper (South) Egypt" [17] the laboratory employed a variety of technologies to confirm paternity, including the STR(short tandem repeats) technology. If the child is male, paternity can be established through chromosomal Y testing. Because the child inherits his father's chromosome. They also mentioned that they may check at 24 different sections of DNA at the same time to confirm paternity, such as (D13S317,). The maternity test, on the other hand, is particularly significant in some circumstances, such as missing children or youngsters on the street, according to the same research. So, if we have a system that is based on a large database of our country, hopefully. We can save all street children.

II. Background , cases and methods

A. Proof of Concept

Firstly, user should login into the system by userName and password then choose the file contains rs numbers, father, mother, child alleles then applying paternity test concepts for Mendelian's law.

There are 8 cases according to dataset for alleles that father have 2 alleles or one allele and same for mother and child so, we have

"'

father mother child alleles (two for each person inherited from both parents)
2 2 2 like : A/A A/A A/A
2 2 1 like : A/A A/G A
2 1 2 like : A/A G A/G
2 1 1 like : A/A G A
1 2 2 like : A G/G A/G
1 2 1 like : G A/T G
1 1 2 like : T G T/G
1 1 1 like : T G T "'

```
//define lists
rsNumberSimilar = []
fatherSimilar = []
motherSimilar = []
childSimilar = []
rsNumbersFather = []
rsNumbersMother = []
fa = []
mo = []
chFather = []
chMother = []
chromosomeFitFather = []
chromosomeNotFitFather = []
chromosomeNotFitMother = []
chromosomeFitMother = []
```

```
//make paternity test function
def pT(father, mother, child, rs, chromosome)
//define list for chromosomes
ch = ['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15',
'16','17','18','19',' 20','21','22','X','Y','MT']
//loop for each element in the list
for f,m,c,r,chro in father,mother,child,rs,chromsome
//case 2 2 2
if chro in ch:
//check that case 2 occured
if len(c) == 2 and len(f) == 2 and len(m) == 2
//check that child have an allele taked from father
if f[0] == c[0] or f[0] == c[1] or f[1] == c[0] or f[1] == c[1]
//push data on lists
rsNumberSimilar.append(r)
fatherSimilar.append(f)
childSimilar.append(c)
chromosomeFitFather.append(chro)
//check that child have an allele taked from mother
if m[0] == c[0] or m[0] == c[1] or m[1] == c[0] or m[1] ==
c[1]
motherSimilar.append(m)
chromosomeFitMother.append(chro)
//check if the child have an allele not taken from father
if f[0] != c[0] and f[0] != c[1] and f[1] != c[0] and f[1] !=
c[1]
rsNumbersFather.append(r)
fa.append(f)
chFather.append(c)
chromosomeNotFitFather.append(chro)
//check if the child have an allele not taken from mother
if m[0] != c[0] and m[0] != c[1] and m[1] != c[0] and m[1]
!= c[1]:
rsNumbersMother.append(r)
mo.append(m)
chMother.append(c)
chromosomeNotFitMother.append(chro)
```

we repeat the same things with each case and push the data on lists like the last example.

we have a function calculate probability that calculate if the child is related to this father or not.

Every get function in the calculate probability return the list.

```
def calculateProbability()
sumSimilar = len(getRsNumberSimilar())
total = (len(getRsNumberSimilar())+
len(getRsNumberFather()))
sumNotSimilar = len(getRsNumberFather())
rule = sumSimilar / total
ruleNotSimilar = sumNotSimilar / total
return rule * 100, ruleNotSimilar * 100
```

B. short tandem repeat

Autosomal short tandem repeat (STR) markers are used in forensic DNA profiling to determine the identify of missing people, validate familial ties, and link people of interest to crime scenes. The algorithm of STR is based on the of count repeats for some nucleotide on DNA ex: ATCATCATCATC should be repeated a lot of times on some specific locations on the DNA (locations are different for each country). The STR counts in the child should be the same as father and mother. [18]

C. Mendel's Laws of Inheritance

The process by which a child inherits genetic information from a parent is known as inheritance. Every family with its rs numbers should have alleles. These alleles in child should be inherited from their father and mother ex: if the child has C/G in a specific rs number then the father should have a C or G in the same rs number and mother. This simply indicates that individuals of the same family share comparable qualities as a result of heredity..

III. RELATED WORK

A. Validation of software for calculating the likelihood ratio for parentage and kinship

J. Dra´bek [1]proposed a method to validate a software for calculating the likelihood ratio in parentage/kinship scenarios when it comes to two programs (paternity index and families). They used in their study seven different test cases in paternity and calculated the error rate and success rate in each of those cases. The seven chosen cases were proved to be thorough but It had some incorrect likelihoods believed to be caused by the software or improper use of the software during the testing phase.

B. R scripts for kinship testing

Masataka [2]T suggested a method using R script for statistical analysis of genetic data for kinship testing. The used methods involve algorithms with R language considering its flexibility with calculations and statistical power and also for conditional probability analysis based on Bayes theorem. Standard paternity trio case and other test cases were conducted and scripts were constructed for each different test case. DNA profiling was used in order to find the allele frequencies of tested loci to establish the links among DNA profiles of individuals

C. PedExpert: a computer program for the application of Bayesian networks to human paternity testing

R R Gomes , S V A Campos S D J Pena [3]proposed a computer program for paternity testing called PedexPert. It is a Windows-based Bayesian network software designed specifically to tackle problems in paternity testing that are complicated due to a lack of genetic information on the claimed father and/or genetic abnormalities. PedexPert enables the development and management of Bayesian networks by implementing algorithms that convert pedigrees and collections of crucial data (genotypes, allele frequencies, and mutation rates) into Bayesian networks. This program can also be used in other cases that include gene mutations or not. The software can create structure of Bayesian networks directly from family pedigrees using Mendelian principles.

D. User-friendly programs for easy calculations in paternity testing and kinship determinations

Wing K. Fung [4]explains and describe the theory and various features of four computer programs which were made for easy calculations of likelihood ratios (LRs) for paternity/kinship determinations based on the calculations of Bayes Theorem, conditional probability and pedigree analysis with the use of computer enumeration to ease the calculations in complex cases in paternity/kinship problems made in each one of the programs. They also mention that these programs can deal with both civil and criminal paternity cases, missing persons and kinship determinations for different disasters such as air-crashes. In each of the different cases that were being tested, they assumed Hardy-Weinberg and linkage equilibrium, which is commonly taken in paternity testing and kinship determinations. The programs were also tested for the standard trio case (mother, a child, and alleged father). They also mention that those programs allow w for 1 or 2 mismatches of DNA profiles due to mutation giving also in their study the computed residual and overall paternity indices as results in each test case.

E. Development of a software for kinship analysis considering linkage and mutation based on a Bayesian network

Chie Morimoto, Hideaki Tsujii, Sho Manabe, Shuntaro Fujimoto, Eriko Hirai, Yuya Hamano, Keiji Tamaki [5]developed a system for kinship analysis based on Bayesian network. The software developed called KinBN and It calculates the likelihood ratios (LRs) at multiple loci considering the effects of linkage and mutation during the testing phase. They compared the results with those of other software with respect to effects of linkage and mutations on the (LRs). They established that their software is an effective tool for kinship analysis and was an improvement when it comes to some other programs.

F. Paternity testing using massively parallel sequencing and the PowerSeq™ AUTO/Y system for short tandem repeat sequencing

Faith [6]proposed a study about a system called an MPS(Massively parallel sequencing(MPS), also known as next-generation sequencing(NGS)) system. (PowerSeq:tm: AUTO/Y) was applied for STR sequencing in the study of first-degree STR sequence allele inheritance from families in Southern Brazil. The results from this study showed advantages of implementing sequence-based analysis, MPS, in paternity testing with improved statistical calculations and a greater resolution for the trios/families tested.

G. A Windows-based software for common paternity and sibling analyses

Jose´ A. Riancho* , Marı´a T. Zarrabeitia [7]proposed a windows based freeware for kinship analysis that is used to calculate likelihood ratios and probabilities of paternity in trio and motherless cases, as well as in cases when a parent is lacking and also compute the probability of two subjects being full-brothers or half-brothers. The software used is called PATCAN which uses a collection of spreadsheets to achieve the goal. Those spreadsheets are used to capture allelic data from the relevant individuals and allelic frequencies which are already stored in the databases. They also mentioned that the software was validated in more than 100 test cases and in all of them the results were compared successfully with those obtained using other packages or by-hand calculations. However, the software PATCAN cannot be used to work out complex cases, such as those with intricate or incomplete pedigrees.

H. Paternity tests in Mexico: Results obtained in 3005 cases

M.E. Garc ıa-Aceves, O. Romero Renter ıa , X.X. D ıaz-Navarro, H. Rangel-Villalobos [8]suggested a study that aimed to describe a posteriori parameters from DNA paternity tests. They reported different results from 3005 pater-nity cases analyzed. Different cases were tested but they found that motherless cases were the most frequent followed by the standard trio cases with the last in place being the kinship reconstruction. The rate that was estimated in the study to which excluded the paternity testing (proved negative) was around 30 percentage in cases due to the STRs loci being mutated.

I. FamLink–a user friendly software for linkage calculations in family genetics

Daniel Kling a,b, *, Thore Egeland b,c , Andreas O. Tillmar d [11]proposed a software called FamLink that can perform statistical calculations on pedigree structures and account for linkage between pairs of differnt markers. Furthermore FamLink can simulate genotype data in order to study the effect of accounting for linkage or not. They used the location of D12S391 on the short arm of chromosome 12. On the other hand, the software can calculate case specific likelihood ratios for two (or more) hypotheses with observed DNA-data for a pair of linked DNA markers and (2) perform simulations for two or more pedigrees (hypotheses) in order to study the impact of ignoring linkage for a specified pair of linked STR markers. The software provides several features which simplifies the interactivity and provides the possibility to include linked markers in relationship calculations.

J. DNA identification of compromised samples with massive parallel sequencing

Andreas Tillmar, Ida Grandell Kerstin Montelius [12]proposed a study that used multiple DNA techniques to identify victims to their relatives. In addition, they used statistical calculations of possibilities of a random match which were achievable since population data from many parts of the world are available. On the other hand, they used the massive parallel sequencing (MPS) which is a technique capable of producing a vast amount of DNA sequence data in a high-through put manner, and panels of single nucleotide polymorphism (SNP) markers which allow the amplification of small DNA fragments, often seen in compromised samples.

K. Beyond simple kinship and identification: a DNA anal- yses from a 17th-19th century crypt in Germany

Amelie Alterauge , Sandra Losch , Andrea Sulzer, Mario Gysi, Cordula [10]proposed a study that were aimed at identifying anticipated and incidental genetic relationships. They used a total of 17 individual's data in their investigation from three families from 17th to 19th centuries. Morphological and radiographic DNA analysis were conducted using the DNA profiles to further study the kinship identification and analysis. The main goal was to test for anticipated and incidental genetic relations and assess the implications of the presence and absence thereof for the assumed identity of the deceased. The research helped to verify or disprove ascribed identities and to elucidate those of hitherto unknown individuals. However, there were some limitations to the kinship analysis when it came to going beyond first to third degree relationships.

L. ForeStatistics: A windows-based feature-rich software program for performing statistics in Forensic DNA analysis, Paternity and relationship testing

Nouman Rasool1*, Waqar Hussain2 [13]proposed a software for performing forensic DNA analysis to establish paternity and relatedness in civil and criminal matters. The software program is called ForeStatistics. It underlines many methods in paternity testing such as DNA statistical calculations, DNA profile management and its matching. The software can estimate random match probabilities from different testing cases of paternity (alleged father, mother child) using different allele frequencies from DNA profiles. The software foreStatisctics were proven to be effective by many researchers in the field for human identification , criminal investigations and aiding the court cases.

M. A Survey of Methods and Tools for Large-Scale DNA Mixture Profiling

Emad Alamoudi, Rashid Mehmood, Aiiad Albeshri, Takashi Gojobori [14]proposed a research in which aims to extend the knowledge of DNA profiling methods and tools with respect to their computational performance and accuracy. They explain how to obtain the DNA profile in forensic science and the factors that contribute to increase the complexity of DNA profile such as number of contributors, population, analytical threshold and other factors that need to be respected in the calculations of DNA profiling. The research also includes many software tools that analyzes the DNA and obtain the DNA profile such as LRmix Studio, TrueAllele, Lab Retriever and other software tools. In conclusion they suggested two software tools (Euroformix and LikeLTD) for DNA profiling given that both utilized most information in the DNA sample and the code for both software tools is available for modification and assessment. However Euroformix has a better GUI.

N. Kinship Determination in Archeological Contexts Through DNA Analysis

Stefania. V, Carlos Eduardo G. A, Martina. L David. C [15]suggested a study to further explore DNA analysis and kinship relations between individuals. Their case studies were the archeological and anthropological analyses of burial sites and skeletal remains. They show using multiple case studies in their research that only genetic analysis can provide a sound and a confirmed determination of kinship. They describe the different molecular strategies for kinship estimation from the classic PCR-based methods to Next Generation Sequencing (NGS).

O. Statistical Softwares Used in Evaluation of Forensic DNA Typing

R. K. Kumawat,Aditi MishraPankaj Shrivastava [16]proposed Statistical analysis and evaluation of an individual's genetic data, as well as interpretation of the results, is ideally necessary to be enhanced by statistical evaluation, especially in mixed DNA profiles produced in sexual assault cases, and numerous software is recommended for managing this.it includes information on software for population data analysis, which is a growing topic in forensic DNA analysis, as indicated by the number of papers on population-based DNA analysis in practically all of the major forensic journals.

PROPOSED APPROACH

P. System overview

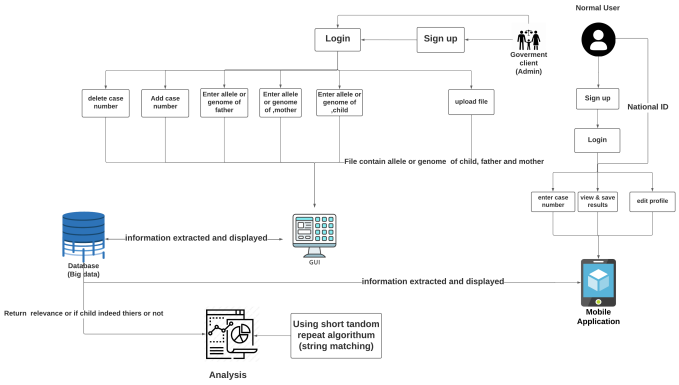
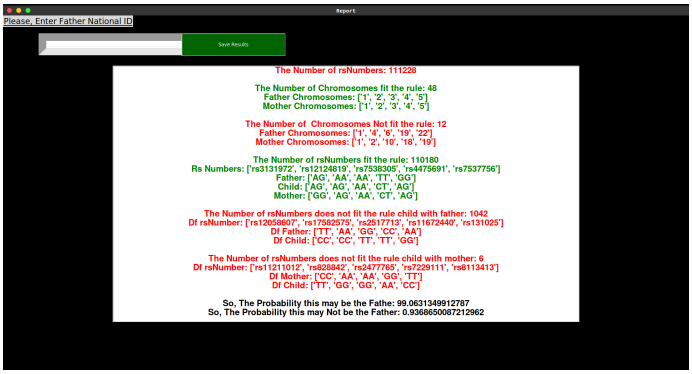


figure (1) : System Overview

The system will be a GUI system that allows some of the competent authorities to upload the data that will be represented on the whole genome. The system will check the similarity between it and each child in our database based on the case presented (different cases for paternity testing) and the system will return to the family the that this child may be related to the family and if there is some relevance. We will build our model depending on the Egyptian genome project because we aim to solve this problem in our countries. In addition, we aim to add another part in our system that proves the parentage of a child to a father depending on rs numbers or the whole genome. In addition, the system will have mobile application the allow users to access ,view and Download the results of specific case by entering their national ID.

IV. EXPERIMENTS



(2) : Report

V. CONCLUSION

Finally, we used the genetics of humans in the best way to deploy our app to help governments in countries to reach the criminal or to know if the child is related to the father that denied their paternity or not. We increased an app that takes the file containing rs numbers, alleles of father, mother, child, and their chromosomes. In addition, the app gives a report to the user about the rs numbers that truly fit the rule of mendelian's law and their chromosomes and what is not fit the rule. On the other hand, the system returns in the report the probability this may be the father or not.

VI. REFERENCES

REFERENCES

[1] Jirı Drabek. "Validation of software for calculating the likelihood ratio for parentage and kinship". In: Forensic Science International: Genetics 3.2(2009), pp. 112–118.

[2] Masataka Takamiya. "R scripts for kinship testing". In: Journal of Iwate Medical Assiociation 73.5 (2021), pp. 189–201

[3] RR Gomes, SV Campos, and SD Pena. "PedExpert: a computer program for the application of Bayesian networks to human paternity testing". In: Genet Mol Res 8 (2009), pp. 273–283.

[4] Wing K Fung. "User-friendly programs for easy calculations in paternity testing and kinship determinations". In: Forensic science international 136.1-3 (2003), pp. 22–34.

[5] Chie Morimoto, Hideaki Tsujii, Sho Manabe, et al. "Development of a software for kinship analysis considering linkage and mutation based on a Bayesian network". In: Forensic Science International: Genetics 47 (2020), p. 102279.

[6] Deborah SBS Silva, Fernanda R Sawitzki, Melissa KR Scheible, et al. "Paternity testing using massively parallel sequencing and the PowerSeq™ AUTO/Y system for short tandem repeat sequencing". In: Electrophoresis 39.21 (2018), pp. 2669–2673.

[7] Jos e A Riancho and Maria T Zarrabeitia. "A Windows-based software for common paternity and sibling analyses". In: Forensic science international 135.3 (2003), pp. 232–234.

[8] ME Garcıa-Aceves, O Romero Renterıa, XX Dıaz-Navarro, et al. "Paternity tests in Mexico: results obtained in 3005 cases". In: Journal of Forensic and Legal Medicine 55 (2018), pp. 1–7.

[9] Luisa Barzon, Enrico Lavezzo, Valentina Militello, et al. "Applications of next-generation sequencing technologies to diagnostic virology". In: International journal of molecular sci- ences 12.11 (2011), pp. 7861–7884.

[10] Amelie Alterauge, Sandra L osch, Andrea Sulzer, et al. "Beyond simple kinship and identification: aDNA analyses from a 17th-19th century crypt in Germany". In: Forensic Science International: Genetics 53 (2021), p. 102498.

[11] Daniel Kling, Thore Egeland, and Andreas O Tillmar. "FamLink–a user friendly software for linkage calculations in family genetics". In: Forensic Science International: Genetics 6.5 (2012), pp. 616–620.

[12] Andreas Tillmar, Ida Grandell, and Kerstin Montelius. "DNA identification of compromised samples with massive parallel sequencing". In: Foren- sic Sciences Research 4.4 (2019), pp. 331–336.

[13] Nouman Rasool and Waqar Hussain. "ForeStatistics: A windows-based feature-rich software program for performing statistics in forensic DNA analysis, paternity and relationship testing". In: Forensic Science International 307 (2020), p. 110142.

[14] mad Alamoudi, Rashid Mehmood, Aiiad Albeshri, et al. "A survey of methods and tools for large-scale dna mixture profiling". In: Smart Infrastructure and Applications. Springer, 2020, pp. 217–248.

[15] Stefania Vai, Carlos Eduardo G Amorim, Martina Lari, et al. "Kinship determination in archeological contexts through DNA analysis". In: Fron- tiers in Ecology and Evolution 8 (2020), p. 83.

[16] RK Kumawat, Aditi Mishra, and Pankaj Shrivastava. "Statistical softwares used in evaluation of forensic DNA typing". In: Forensic DNA Typ- ing: Principles, Applications and Advancements. Springer, 2020, pp. 105– 134.

[17] Ghada Ali Farghali Omran. Forensic DNA profiling of a population sample from upper (south) Egypt. University of Leicester (United Kingdom), 2008.

[18] Wyner, Nicole, Mark Barash, and Dennis McNevin. "Forensic autosomal short tandem repeats and their potential association with phenotype." Frontiers in genetics (2020): 884.