Eksakta : Berkala Ilmiah Bidang MIPA
VOLUME 21 NO 01 2020, pp 46-53
ISSN : Print 1411-3724 — Online 2549-7464
DOI : https://doi.org/10.24036//eksakta/vol21-iss1/216

Eksakta
Berkala Ilmiah Bidang MIPA
http://www.eksakta.ppj.unp.ac.id/index.php/eksakta

Article

The Similarity Measurement of Human DNA Profile Using Fuzzy Similarity

Article Info

Article history :

Received April 13, 2020
Revised April 20, 2020
Accepted April 22, 2020
Published April 30, 2020

Keywords :

human DNA profile,
fuzzy similarity measure,
biological family

Meira Parma Dewi[*1], Nurtami Soedarsono[2]

[1]Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Negeri Padang, Padang, Indonesia
[2]Department of Basic Dentistry, Faculty of Dentistry, Universitas Indonesia, Depok, Indonesia

Abstract. This research investigated the similarity of human DNA profile using fuzzy similarity measure. The similarity measurement of DNA profile had been done by measuring the similarity between query's DNA profile and its biological family such as father, mother, brother, sister, grandmother and grandfather. The similarity measurement had been done to the short tandem repeat (STR) alleles in sixteen loci. The result of the experiment showed that each simulation gave matching result. This research is useful for Indonesian National Police (POLRI) in identifying process of disaster victim, terrorism victim and other criminal conduct.

This is an open acces article under the CC-BY license.

This is an open access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2017 by author.

Corresponding Author :
Meira Parma Dewi
Department of Mathematics, Faculty of Mathematics and Natural Science (FMIPA)
Universitas Negeri Padang, Padang, Indonesia
Email : meiraparma@fmipa.unp.ac.id

1. Introduction

Indonesia is a country that is vulnerable to natural disasters such as floods, earthquakes, tsunamis and so on[1-8]. In addition, in recent years Indonesia has also become a country that has become one of the targets of acts of terrorism. Many terrorism and natural disasters have caused

many fatalities, some of which can no longer be recognized or identified due to damage to some or all parts of the victim's body. To be able to identify victims, genetic information is needed (DNA (deoxyrebous nucleid acid) from the victim[9-20].

DNA is a biological fingerprint of an individual obtained from the genetic inheritance of parents [9-20] (21, 22, 24), DNA profile identification is carried out on DNA evidence contained in human body tissues such as hair, teeth, blood, urine, saliva, muscles and sperm (23). The DNA profile maps the genetic makeup of individuals consisting of 16 loci, namely CSF1PO, D13S317, D16S539, D18S51, D19S433, D21S11, D2S1338, D3S1358, D5S818, D7S720, D8S1179, FGA, TH01, T19X, D21S11, D2S1338, D3S1358, D5S818, D7S720, D8S1179, FGA, TH01, T19X, and D21S11, D2S1338, D3S1358, D5S818, D7S720, D8S1179, FGA, TH01, TPOX, and VWA. A locus has two alleles that indicate the value of short tandem repeat (STR) which is an integer. However, due to several factors the STR value of the locus allele can shift 0-1 to the right or 0-1 to the left(25-27).

The identity of the victims can be investigated, DNA comparison profile data from the biological families of the victims are needed. Measurement of the similarity of the profile of human DNA can be done by comparing the STR of each locus. In comparing the two alleles the STR value must be exactly the same, but due to the possibility of shifting the STR value if the STR values being compared are not really the same it will be considered mismatched. To be able to measure the similarity of two prfile DNA alleles that have shifted used fuzzy similarity measurement.

Experimental Section

Fuzzy Inference System

Measurement of the similarity of DNA profiles will be done by comparing each allele at 16 loci of the DNA profile of the victim with a DNA profile from the reference to the alleged biological family of the victim. In this case the family used as a comparison of the DNA profile is the father, mother, grandfather and grandmother on the part of the father and mother. Grandparents will be used as a reference if one or both parents are not present.

The rules in matching the DNA profile of the victim with the DNA profile of the father and mother are

If [(allele _1_ victim ≈ allele _1_ father) OR (allele _1_ victim ≈ allele _2_ father)] AND [(allele _2_ victim ≈ allele _1_ Mother) OR (allele _2_ victim 2 allele _2_ mother)] THEN matches

If [(allele _1_ victim ≈ allele _1_ mother) OR (allele _1_ victim ≈ allele _2_ mother)] AND [[allele _2_ victim ≈ allele _1_ Father) OR (allele _2_ victim 2 allele _2_ Father)] THEN matches

If the father's DNA profile data is not available then in the process of measuring the similarity the DNA profile will use the DNA profile of the father and biological mother of the victim's father as a substitute for the victim's father. the process of measuring the resemblance of a locus does not change, where one locus must match / similar to the father's reference (ref_A) and the other must match / similar to the reference from the mother's party (Reff_B).

If [(allele _1_ victim ≈ allele _1_ Reff_A) OR (allele _1_ victim ≈ allele _2_ Reff_A)] AND [(allele _2_ victim ≈ allele _1_ Reff_B) OR (allele _2_ victim ≈ allele _2_ Reff_B)] THEN matches

If [(allele _1_ victim ≈ allele _1_ Reff _B) OR (allele _1_ victim ≈ allele _2_ Reff_B)] AND [[allele _2_ victim ≈ allele _1_ Reff_A) OR (allele _2_ victim ≈ allele _2_ Reff_A)] THEN matches

Membership Functions of Input Variables and Output Variables

Input variables have three membership functions, namely small, medium and large.

$$f_{kecil}(x) = \begin{cases} 1 & x \le 0.15 \\ \dfrac{0.3 - x}{0.15} & 0.15 \le x \le 0.3 \\ 0 & x \ge 0.3 \end{cases}$$

$$f_{sedang}(x) = \begin{cases} 0 & x \le 0.2; x \ge 0.5 \\ \dfrac{x - 0.2}{0.15} & 0.2 \le x \le 0.35 \\ \dfrac{0.5 - x}{0.15} & 0.35 \le x \le 0.5 \end{cases}$$

$$f_{besar}(x) = \begin{cases} 0 & x \le 0.4 \\ \dfrac{x - 0.4}{0.1} & 0.4 \le x \le 0.5 \\ 1 & x \ge 0.5 \end{cases}$$

This membership function is described as follows:

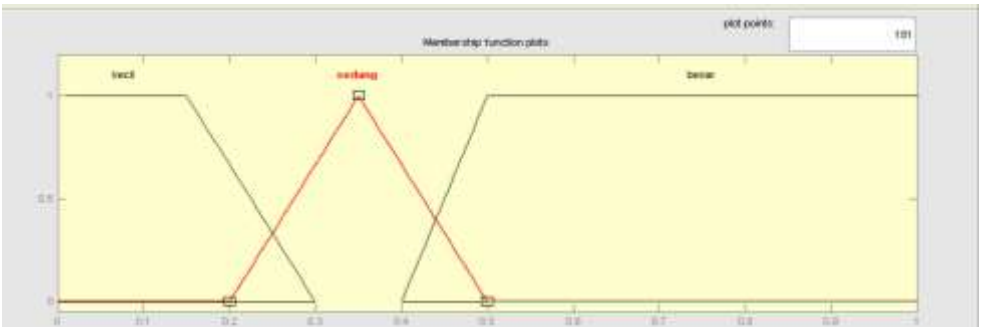

Figure 1. Input Variable Membership Function

The output variable has three membership functions, namely low value 0, medium value 0.5 and high value 1.

Fuzzy rules used are as follows:

If (father is small) And (mother is small) then (similarity is low)
If (father is small) And (mother is medium) then (similarity is low)
If (father is small) And (mother is big) then (similarity is medium)
If (father is medium) And (mother is small) then (similarity is low)
If (father is medium) And (mother is medium) then (similarity is medium)
If (father is medium) And (mother is big) then (similarity is high)

If (father is big) And (mother is small) then (similarity is medium)
If (father is big) And (mother is medium) then (similarity is high)
If (dad is big) And (mom is big) then (similarity is high)
If (mother is small) And (grandfather is not big) And (grandmother is Small)
Then (similarity is low)
If (mother is small) And (grandfather is big) And (grandmother is Small)
Then (similarity is medium)
If (mother is small) And (grandfather is small) And (grandma is big)
Then (similarity is medium)
If (mother is small) And (grandfather is small) And (grandma is not big)
Then (similarity is low)
If (mother is medium) And (grandfather is not big) And (grandmother is Small)
Then (similarity is low)
If (mother is medium) And (grandfather is big) And (grandmother is Small)
Then (similarity is high)
If (mother is medium) And (grandfather is small) And (grandmother is large)
Then (similarity is high)
If (mother is medium) And (grandfather is small) And (grandma is not big) Then (similarity is medium)
If (mother is big) And (grandfather is not big) And (grandmother is Small) Then (similarity is medium)
If (mother is big) And (grandfather is big) And (grandmother is Small) Then (similarity is high)
If (mother is big) And (grandfather is small) And (grandma is big) Then (similarity is high)
If (mother is big) And (grandfather is small) And (grandmother is not big) Then (similarity is medium)
If (father is small) And (grandfather of mother is not big) And (grandmother of mother is small) Then (similarity is low)
If (father is small) And (grandfather of mother is big) And (grandmother of mother is small) Then (similarity is medium)
If (father is small) And (grandfather of mother is small) And (grandmother of mother is large) Then (similarity is medium)
If (father is small) And (grandfather of mother is small) And (grandmother of mother is not big) Then (similarity is low)
If (father is medium) And (grandfather of mother is not big) And (grandmother of mother is small) Then (similarity is low)
If (father is medium) And (grandfather of mother is large) And (grandmother of mother is small) Then (similarity is high)
If (father is medium) And (grandfather of mother is small) And (grandmother of mother is large) Then (similarity is high)
If (father is medium) And (grandfather of mother is small) And (grandmother of mother is not big) Then (similarity is medium)
If (father is big) And (grandfather of mother is not big) And (grandmother of mother is small) Then (similarity is medium)
If (father is big) And (grandfather of mother is big) And (grandmother of mother is small) Then (similarity is high)
If (father is big) And (grandfather of mother is small) And (grandmother of mother is big) Then (similarity is high)
If (father is big) And (grandfather of mother is small) And (grandmother of mother is not big) Then (similarity is medium)

If (grandfather is not big) and (grandmother is not big) and (grandfather of mother is not big) and (grandmother of mother is not big) then (similarity is low)
If (grandfather is big) and (grandmother is not big) and (grandfather of mother is not big) and (grandmother of mother is not big) then (similarity is medium)
If (grandfather is not big) and (grandmother is big) and (grandfather of mother is not big) and (grandmother of mother is not big) then (similarity is medium)
If (grandfather is not big) and (grandmother is not big) and (grandfather of mother is big) and (grandmother of mother is not big) then (similarity is medium)
If (grandfather is not big) and (grandmother is not big) and (grandfather of mother is not big) and (grandmother of mother is big) then (similarity is medium)
If (grandfather is big) and (grandmother is not big) and (grandfather of mother is big) and (grandmother of mother is not big) then (similarity is high)
If (grandfather is big) and (grandmother is not big) and (grandfather of mother is not big) and (grandmother of mother is big) then (similarity is high)
If (grandfather is not big) and (grandmother is big) and (grandfather of mother is big) and (grandmother of mother is not big) then (similarity is high)
If (grandfather is not big) and (grandmother is big) and (grandfather of mother is not big) and (grandmother of mother is big) then (similarity is high)

Measurement of Allel Similarity

Fuzzy similarity measurements of DNA profiles are done by measuring the similarity of an allele. Assuming that a triangular allele with a short tendem repeat (STR) of an allele shows the middle value, the distance of the two legs is the same ie 0.4 and the height of the allele is equal to 1. So to measure the similarity of the alleles compared are used the equation:

$$\frac{t}{1} = \frac{\frac{1}{2}(a3 - b1)}{(a3 - a2)}$$

Where: first allele position <second allele
t = the intersection point of the two alleles
a2 = STR value from the first allele
a3 = a2 + 0.2
b1 = second allele STR value - 0.2
So $0 \leq t \leq 1$

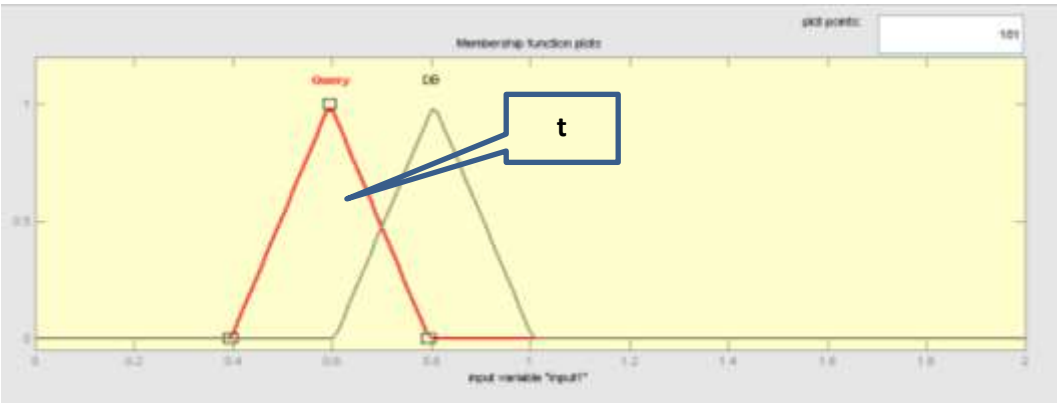

Figure 2. Measurement of the Similarity of Two Alleles

Results and Discussion

The data used as input for the system is a complete DNA profile data consisting of 16 loci, each consisting of two alleles. To enter DNA profile data into the system, it is done manually. Data obtained from the identification of biological evidence (DNA evidence) by a PCR machine in the form of an electropherogram still contains noise. This noise is not taken into account in determining a person's DNA profile. So for each DNA profile loci on the electropherogram only the two highest signals are read. This signal shows the alleles from the relevant loci. If there are two high signals in a loci, then both are alleles, but if there is only one signal that is significantly high compared to the surrounding noise, the first allele and the second allele for the relevant loci have the same value.

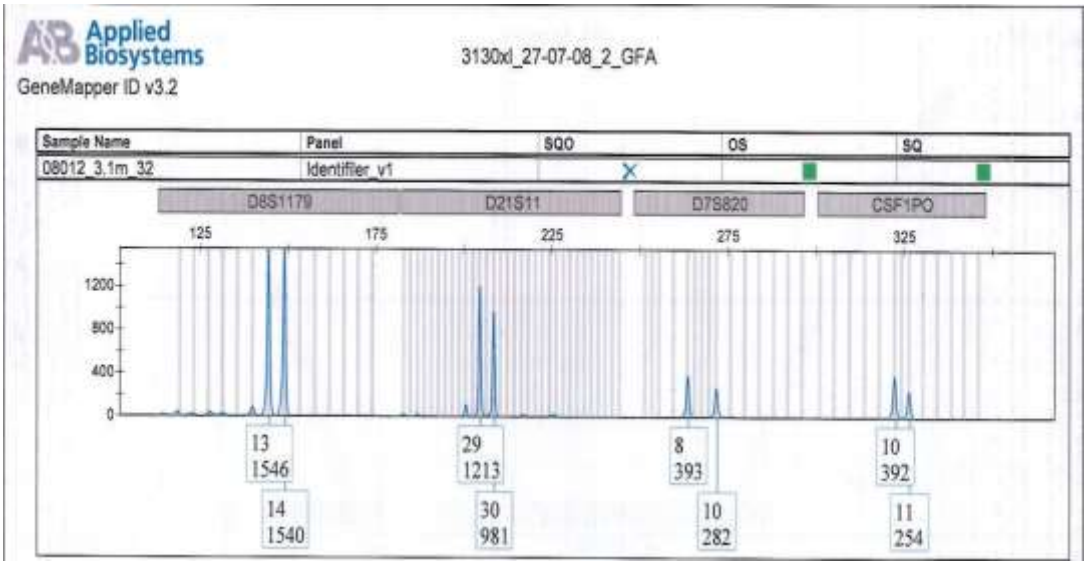

Figure 3. Electropherogram Signal Indicating an Allele from the Locus

Alleles are represented as an isosceles triangle, where the distance between the two legs is 0.4 and height is 1, the midpoint of the two feet is the STR value indicated by an allele. For a loci, each allele will be measured for its fuzzy similarity to the alleles that are in reference to the same loci. Similarity value is obtained by adding up the similarity value of each loci divided by 32.

Measurement of the similarity of DNA profiles using fuzzy similarity measurements is done by giving a similarity value to each allele which then produces a similarity value from a locus. The average of the similarity values of all loci is the similarity value of the DNA profile. DNA profile matches can be said to match if similarity values> 0.5.

Conclusions

DNA profile is a biological fingerprint that is owned by every human being that can distinguish the identity of an individual with other individuals. To facilitate the process of identifying victims of disaster, DNA reference profiles from the biological families of victims are needed. Fuzzy similarity measurement is used in the process of measuring the similarity of a human DNA profile because the alleles shown at the DNA profile locus can experience a shift caused by several factors. If the

alleles shift in the range of 0.2 to the right or 0.2 to the left, a similarity value of 0.5 will be obtained so that the two alleles being compared can be said to be suitable or similar. To conclude a person's DNA profile with a reference said to be similar then the value of each allele from each DNA profile locus is summed then divided by 32, so that a similarity value> = 0.5 is obtained.

5. Acknowledgement

Thank you to Almarhum Dr. M. Rahmat Widyanto, M. Eng who provided a lot of help and contributions. This research was part of my master's research while studying for a master's degree at the University of Indonesia.

References

[1] Rai NK, Rim KI, Wulandari EW, Subrata F, Sugihantono A, Sitohang V.(2020). Strengthening emergency preparedness and response systems: experience from Indonesia. WHO South-East Asia journal of public health;9:26-31.

[2] Jang S, Ekyalongo Y, Kim H.(2020). Systematic Review of Displacement and Health Impact From Natural Disasters in Southeast Asia. Disaster medicine and public health preparedness:1-10.

[3] Meilasari-Sugiana A, Endro G.(2019). Shaping collective action for community-based disaster management in Merapi, Central Java, Indonesia. Journal of emergency management;17:385-401.

[4] Vesterinen HM, Dutcher TV, Errecaborde KM, Mahero MW, Macy KW, Prasarnphanich OO, et al.(2019). Strengthening multi-sectoral collaboration on critical health issues: One Health Systems Mapping and Analysis Resource Toolkit (OH-SMART) for operationalizing One Health. PloS one;14:e0219197.

[5] Yoshida M, Yanuaryska RD, Shantiningsih RR, Mudjosemedi M, Honda E.(2019). Comparison of radiation risk perception and knowledge of radiation between Indonesian and Japanese dental students. Journal of environmental radioactivity;204:104-10.

[6] Abbas Khan K, Zaman K, Shoukry AM, Sharkawy A, Gani S, Sasmoko, et al.(2019). Natural disasters and economic losses: controlling external migration, energy and environmental resources, water demand, and financial development for global prosperity. Environmental science and pollution research international;26:14287-99.

[7] Elsi M, Novera I.(2019). Different triage categorization using Emergency Severity Index (ESI) method in emergency department. Enfermeria clinica;29 Suppl 1:101-4.

[8] Yusvirazi L, Ramlan AAW, Hou PC.(2018). State of emergency medicine in Indonesia. Emergency medicine Australasia : EMA;30:820-6.

[9] Alotaibi SS, Sayed SM, Alosaimi M, Alharthi R, Banjar A, Abdulqader N, et al.(2020). Pollen molecular biology: Applications in the forensic palynology and future prospects: A review. Saudi journal of biological sciences;27:1185-90.

[10] Tytgat O, Gansemans Y, Weymaere J, Rubben K, Deforce D, Van Nieuwerburgh F.(2020). Nanopore Sequencing of a Forensic STR Multiplex Reveals Loci Suitable for Single-Contributor STR Profiling. Genes;11.

[11] Sahoo S, Samal R, Behera S, Swain AK, Biswas S, Shrivastava P, et al.(2020). Genomic portrait of Odisha, India drawn by using 21 autosomal STR markers. International journal of legal medicine.

[12] Szkuta B, Ansell R, Boiso L, Connolly E, Kloosterman AD, Kokshoorn B, et al.(2020). DNA transfer to worn upper garments during different activities and contacts: An inter-laboratory study. Forensic science international Genetics;46:102268.

[13] Karadayi S, Moshfeghi E, Arasoglu T, Karadayi B.(2020). Evaluating the persistence of laundered semen stains on fabric using a forensic light source system, prostate-specific antigen Semiquant test and DNA recovery-profiling. Medicine, science, and the law;60:122-30.
[14] Young JM, Linacre A.(2020). Use of a Spray Device to Locate Touch DNA on Casework Samples. Journal of forensic sciences.
[15] Martin B, Kanokwongnuwut P, Taylor D, Kirkbride KP, Armitt D, Linacre A.(2020). Successful STR amplification of post-blast IED samples by fluorescent visualisation and direct PCR. Forensic science international Genetics;46:102256.
[16] Assenmacher DM, Fields SD, Crupper SS.(2020). Comparison of Commercial Kits for Recovery and Analysis of Bacterial DNA From Fingerprints. Journal of forensic sciences.
[17] Scott L, Finley SJ, Watson C, Javan GT.(2020). Life and death: A systematic comparison of antemortem and postmortem gene expression. Gene;731:144349.
[18] Nakamura M, Idota N, Shintani-Ishida K, Hitosugi M, Ikegaya H.(2020). Simple, Frequent Indicator for Personal Identification-Postmortem and Antemortem Abdominal Computed Tomography Findings of a Charred Body. The American journal of forensic medicine and pathology;41:56-9.
[19] Goray M, Kokshoorn B, Steensma K, Szkuta B, van Oorschot RAH.(2020). DNA detection of a temporary and original user of an office space. Forensic science international Genetics;44:102203.
[20] Phan K, Barash M, Spindler X, Gunn P, Roux C.(2020). Retrieving forensic information about the donor through bacterial profiling. International journal of legal medicine;134:21-9.
[21] Troubleshooting for PCR and multiplex PCR. http://www.med.yale.edu/genetics/ward/tavi/Trblesht.html
[22] DNA Profiling In Forensic Science http://www.nzic.org.nz/ChemProcesses/biotech/12D.pdf
[23] Capillary Electrophoresis, http://www.beckman.com/resourcecenter/labresources/ce/cedefinitionmodes.asp
[24] Parma Dewi, Meira, (2009) Perancangan Model Pengukuran Kemiripan Profil DNA Manusia Menggunakan Ukuran Kemiripan Fuzzy, Universitas Indonesia.
[25] Widyanto MR, Hartono Raggio, Soedarsono, Nurtami. (2016). A Novel Human STR Similarity Method Using Cascade Statistical Fuzzy Rule With Tribal Inforation Inference, International Journal of Electrical and Computer Engineering vol.6 Desember
[26] Anggreainy, Maria Susan et al. (2018). Gaussian Fuzzy Number For STR-DNASimilarity Calculation Involving Familial And Tribal Relationships, Hindawi Andvance in Bioinformatics vol 2018.
[27] Taoufik G., Benslimean, Daoudi. (2006). "Fuzzy Similarity Measure", Springer.