

令和元年6月10日現在

機関番号：14301

研究種目：基盤研究(B) (一般)

研究期間：2016～2018

課題番号：16H05273

研究課題名(和文) 包括的遺伝情報プロフィール創生に向けた分析技術と数学的解釈の実践的融合

研究課題名(英文) Practical merging DNA typing technology and mathematical interpretation of the typing results towards creation of comprehensive forensic DNA profiles from mixture samples

研究代表者

玉木 敬二 (TAMAKI, Keiji)

京都大学・医学研究科・教授

研究者番号：90217175

交付決定額(研究期間全体)：(直接経費) 11,700,000円

研究成果の概要(和文)：法医鑑識領域では複数人のDNAが混合している試料の解析が急増している。本研究では、キャピラリー電気泳動法(CE法)によるDNA検査結果を、continuous modelを原理として数学的解釈を行うソフトウェア(Kongoh)の開発・検証を行い、問題となる人の試料の関与について高い確率を示すことができた。また、次世代シーケンシング(NGS)の法科学応用として、実験的混合試料の分析を試みた。その結果、ローカス毎に得られるリード数は大きく異なったが、全体的な検出感度はCE法に比べて上がった。しかし、目的とするアリルを正しく抽出できない場合もみられたためソフトウェアの改良などの検討が必要と考えられる。

研究成果の学術的意義や社会的意義

本研究では急増する複数人のDNAが混合した試料の検査結果の数学的分析手法として、閾値設定を必要としないコンティニューアモデルに基づいた新しいソフトウェアを作成し、実験的混合試料などによって検証した。このソフトウェアは自動的に最適な関与人数を推定するというこれまでにない特徴があり、プログラムに馴染みのない人でも使用することができる。また、ウェブ上にソースコードが公開されているので、アルゴリズムの検証ができる。また、次世代シーケンシングを混合試料の解析に初めて試用した。その結果、検出感度は従来法より高い反面、実務応用には課題が残されていることがわかり、今後の研究の方向性を示すことができた。

研究成果の概要(英文)：In forensic casework, the number of DNA analysis of mixture samples of two or more individuals has rapidly increased. We developed a new software (named as "Kongoh") which can estimate the likelihood ratio (LR) about contribution of a person of interest (POI) by analyzing DNA typing results from the capillary electrophoresis (CE) based on a continuous model, and validated it by referring to the SWGDAM guidelines. Kongoh showed much higher LR values about a POI's contribution with those based on other models. Analysis of experimentally generated DNA mixture samples by the high-throughput sequencing (NGS) were also performed whether it can be applied to forensic mixture analysis. The sensitivity of NGS was higher than that of conventional CE although the number of NGS reads differs depending on the locus, and a few loci could not obtain any reads. It is necessary to revise the plug-in software which extracts the exact target STR sequences to improve the accuracy of mixture analysis.

研究分野：法医学、法遺伝学

キーワード：DNA多型 マイクロサテライト STR 微量混合試料 continuous model

1. 研究開始当初の背景

(1)混合試料分析に苦慮する DNA 鑑定

法医学鑑識領域で行われる DNA 鑑定はゲノム領域にある繰り返し配列 (STR) を PCR によって増幅して、キャピラリー電気泳動法 (CE 法) によって検出して型判定する方法が用いられている。また、昨今の犯罪の巧妙化などのため、現場に単独のヒト DNA 試料が殆ど残らなくなり、あっても被害者などと混合したものが多く、このため、急増する複数のヒト由来の DNA が混合した試料は実務 DNA 検査の中で現在最も重要な課題となっているが、わが国では分析が全く行われていない。さらに、DNA 分析技術の発展は微量の DNA も検出するため、アリのドロップアウトが生じる stochastic effect (確率効果、SE) が起こったり、アリの 1 リピート前に現れる PCR 副産物 (スタター) と微量 DNA 由来の真のピークの判別が困難となっている。

(2)混合試料の解釈方法の発展

混合試料の問題を解決するために、binary model (BM) や semi-continuous model (SCM) による分析法が考案されたが、アリかスタターかの判断は人為的であり、解釈の客観性は損なわれている。この問題を解消する方法として考案されたのが continuous model (CM)¹⁾ 法である。検出されたアリピーク高は、ローカスの増幅効率、スタター、PCR 断片長に応じた減衰率などの多くの因子によって変容している。CM は混合試料における関与人数、関与者の様々な STR 型の組み合わせと DNA 量比を想定して、前記因子をパラメーターとして、予想されるピークのピーク高を deconvolution (関数によってもとの原信号を引き出す操作) して、観察されたチャートのピーク高とのずれの程度を重み (ウェイト) として表す。問題となる人 (person of interest、POI) の DNA が含まれている場合と、含まれていない場合について現場試料の型となる尤度を求め、その比 (尤度比) を算出して POI の関与についての判断をする。CM はスタターの閾値や関与者の可能な組合せを設定する必要がない点で、それまでの方法に比べ、関与者の候補となる STR 型の組み合わせに漏れがなく、より客観的で正確な分析が可能となった。

(3)次世代シーケンサーの利点と課題

次世代シーケンサー (NGS) が DNA 研究に利用されるようになったが、法医学鑑識分野にとっても活用が期待され²⁾、STR 型検査専用のパネルがいくつか市販されている。NGS 法による STR 型検査では、STR 領域の繰り返し回数だけでなくシーケンス情報も得ることができるため、CE 法に比べてアリの数が増加する³⁾。これにより、混合試料に含まれている人数や混合比率を CE 法よりも高い精度で推定できると期待される。しかし、NGS 法による混合試料の解釈については未だ十分な検討がなされていない。正確な混合試料の解釈を行うためには、リード数のばらつきやアリの検出率、アーティファクトやエラーの発生率などの基礎的なデータの検討が不可欠であるが、十分なデータが収集できていないのが現状である。

2. 研究の目的

現在の物体検査の主流となっている混合試料の DNA 検査結果の精度の高い数学的解釈が急務となっているが、わが国では全く手つかずの状態である。本研究では検出閾値を決めずピーク高を利用して

分析する continuous model (CM) に基づくソフトウェアの作成を目指す。また、作成されたソフトウェアについて、米国のガイドラインに沿って検証を行い、その精度を保証する。ソフトウェアは Web 上で公開し、他の研究者の評価や利用を促進する。一方、次世代シーケンサー (NGS) は膨大な DNA 多型情報をもたらすため、法医鑑識領域にとって究極的な活用法といえるが、同時に配列決定の精度など実務応用に関する課題も残されている。本研究では試験的作成混合試料の STR タイピングに NGS を用いることにより、高い精度が要求される混合試料分析にどのように応用できるのかを検討する。

3. 研究の方法

(1) 新しい STR 検査システムにおける混合試料の数学的評価

日本人の GlobalFiler 検査 (GF 検査) の 21 STR のアリル頻度⁴⁾を基に、コンピュータ上で模擬の STR 型 10 万人分、及び 2 人から 5 人までの混合試料を各 10 万組作成した。模擬の STR 型 10 万人分を利用して、RMP を算出した。各混合試料については、アリルピーク高を考慮しない簡便な方法 (binary model) に基づき、関与人数の推定と LR の算出を行った。一連の解析は R 言語 で作成したプログラムを利用して行った。

(2) コンティニューアモデルに基づいた混合微量試料解析ソフトウェアの作成と検証

CM 法による尤度比算出の実際のアルゴリズムを構築した。一連のプログラム作成は R 言語で作成したプログラムを利用して行った。また、R 言語に馴染みのない研究者にも利用が可能なように、graphical user interface を用いて、アクセスしやすい工夫をした。

完成されたソフトウェアの解析能力の検証を、米国の SWGDAM のガイドライン⁵⁾に従いおこなった。具体的には、感度、特異度、再現性、計算の正確性、模擬事例検討について、計算を行い検証した。また、完成したソフトウェアについて、まず、2-4 人の試験的に作成した混合 DNA 試料を AmpF₀STR® Identifiler® Plus PCR Amplification Kit (IDP) にて STR 型判定を行った結果を基に、BM 法や SCM 法 (*LRmix Studio*)⁶⁾を用いた結果と比較し、その精度を検討した。さらに、CM 法の公開ソフトウェアである *EuroForMix*⁷⁾や *likeLTD*⁸⁾を用いて解釈結果を比較した。

(3) 単一試料および混合試料における NGS 法による STR ローカスのタイピング

単一個人 5 試料と混合試料 10 試料について、Precision ID GlobalFiler™ NGS STR Panel v2 を使用して GlobalFiler Panel に含まれる常染色体 31 STR ローカスを対象とした。各試料について、総 DNA 量は 1 ng とした。ライブラリの調製には Precision ID Library Kit を使用し、Ion Chef™ Instrument によりテンプレート調製を行った。また、シーケンシングは Ion PGM™ System により行った。各ステップでの実験条件は Precision ID Panels with Ion PGM™ System APPLICATION GUIDE に従った。また、CE 法として各試料の DNA 1ng から IDP にて STR 型判定を行った。検出閾値は 150 RFU とした。ローカスごとの総リード数、アリルの検出率、エラーの発生率について NGS 法と CE 法を比較検討した。

4. 研究成果

(1)GF 検査の混合試料における識別能力

GF 検査の random match probability の中央値は 3.12×10^{-28} となり、IDP 検査の場合 (2.10×10^{-19}) よりも個人識別力が向上した。推定関与人数の正答率は、2 人の混合試料で 99.97%、3 人で 97.3%、4 人で 87.4% となり、IDP 検査の場合 (2 人:99.5%、3 人:90.9%、4 人:74.9%) よりも高い正答率を示した。また、被疑者の関与に関する LR の値が 10,000 以上となる場合は、2 人の混合試料で全例、3 人で 99.995%、4 人で 98.9%、5 人で 87.8% であった。よって、GF 検査は混合試料の数学的評価に有用といえる。

(2)混合微量試料解析ソフトウェア”Kongoh”の作成と精度検証

CE 法における continuous model によるソフトウェア”Kongoh”を完成させ、その精度について実験的混合試料により検証した。2 人から 4 人までの DNA を様々な量や比率で混合した計 72 試料を作製してマイクロサテライト 15 領域の DNA 型のタイピングを行い、検査結果を作成したソフトウェアで解析した。その結果、2 人、3 人の混合試料では DNA 量が微量であっても何人の DNA が含まれるかを高い精度で推定できた。また、実際に試料中に DNA を含む個人では正しくその混入を支持するような確率が得られ、DNA を含まない個人では混入を支持しない結果となった。ただし、4 人の混合試料で微量試料の場合何人の DNA が含まれるか、ある個人が DNA を含むか否かの判断が難しいことも示した。よって、ソフトウェアは高い精度を有しているが、15 ローカスの IDP 検査は 3 人の混合試料までが識別の限界であることを示唆し、2019 年 4 月から導入された 21 ローカスの GF 検査の活用が期待される。

また、Kongoh の解析能力の検証を行ったところ、Kongoh は他の CM 法に基づくソフトウェアとは異なり、自動的に最適な関与者人数を判断し、尤度比を算出する特徴があるなど、精度の高いソフトウェアであることが示された。Kongoh は他の研究者の利用をはかるため、Web (GitHub) に挙げられ、無料のライセンス利用が付加された。このため、他機関での試用も行われた。現在、GF 検査に対応するため、パラメーターの値などを実験データより導入してソフトウェアの大幅な変更を行うため、実験データの解析を進めている。

(3)混合試料分析における NGS 法の評価

単一個人試料ではローカス間で総リード数に大きな差がみられた。特に、最も多いリード数を示した TH01 と最も少ないリード数を示した Penta_D の間では 32 倍の差があった。また、混合試料について、NGS 法および CE 法にて drop-out が起こったアレルの数は、2 名の混合試料では混合比が 9:1 であれば drop-out は minor contributor (mc) 由来の 1 つのアレルに留まっていたが、混合比が 19:1 の場合には 7 個の drop-out が発生した。しかし、NGS 法では個人あたりの DNA 量が 100 pg 以上であればほとんど drop-out は発生しないことが分かった。一方、CE 法では、試料中の mc が単独でもつアレルはすべて drop-out しており、NGS 法が CE 法よりもアレルの検出率は高いと考えられる。

しかし、十分量の DNA がある試料の分析でも目的とする STR シーケンスが抽出されない“エラー”と思われる配列が観察された。また、混合試料中の各個人の DNA 量の差が大きくなるほど、このエラーは増える傾向にあった。これらの塩基配列を確認したところ、①STR の繰り返し配列の前後に 1 塩基の挿入

があるもの、② STR の繰り返し配列の中で 1 塩基の挿入、欠損あるいは置換があるもの、③通常の STR 領域の塩基長に比べて極端に短いものが確認された。

<引用文献>

1) Perlin MW, et al. J Forensic Sci, 56: 1430-47, 2011. 2) Børsting C, Morling N. Forensic Sci Int Genet, 18: 78-89, 2015. 3) Gettings KB, et al. Forensic Sci Int Genet, 21: 15-21, 2016. 4) Fujii K, et al. Legal Med, 17: 306-8, 2015. 5) Scientific Working Group on DNA Analysis Methods. https://media.wix.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf. 2015. 6) Haned H, et al. Forensic Sci Int Genet. 6: 762-74, 2012. 7) Bleka Ø, et al. Forensic Sci Int Genet. 21: 35-44, 2016. 8) Steele CD, et al. Stat Appl Genet Mol Biol. 15: 431-45, 2016.

5. 主な発表論文等

〔雑誌論文〕 (計 6 件)

① 真鍋翔, 森本千恵, 藤本駿太郎, 濱野悠也, 宮尾昌, 玉木敬二. 複数回の DNA 検査結果を統合した尤度比算出法. DNA 多型. 査読有. 2018; 26(1): 125-129.

② Manabe S, Morimoto C, Hamano Y, Fujimoto S, Tamaki K. Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model. PLOS ONE. 査読有. 2017; 12(11): e0188183.

DOI: 10.1371/journal.pone.0188183

③ Manabe S, Hamano Y, Morimoto C, Kawai C, Fujimoto S, Tamaki K. New stutter ratio distribution for DNA mixture interpretation based on a continuous model. Leg Med (Tokyo). 査読有. 2016; 19: 16 -21.

<https://doi.org/10.1016/j.legalmed.2016.01.007>

〔学会発表〕 (計 1 1 件)

① 平井瑛里子, 橋谷田真樹, 森本千恵, 藤本駿太郎, 濱野悠也, 尾関宗孝, 玉木敬二. Precision ID GlobalFiler™ NGS STR Panel を利用した次世代シーケンスによる混合試料解析の検討. 日本 DNA 多型学会第 27 回学術集会. 講演要旨集. 2018; p70, 松江.

② 真鍋 翔, 濱保英樹, 小谷泰一, 宮尾 昌, ジェマイル レイラ, 玉木敬二. GlobalFiler 21 ローカスにおける DNA 型の一致率と混合試料の数学的評価について. 日本 DNA 多型学会第 27 回学術集会. 講演要旨集. 2018; p104, 松江.

③ Manabe S, Morimoto C, Hamano Y, Hirai E, Fujimoto S, Tamaki K. Comparison of three software for DNA mixture interpretation based on the continuous model. The 29th International Symposium on Human Identification. Abstracts. 2018; p66, Phoenix, United States of America.

〔図書〕 (計 0 件)

〔産業財産権〕

○出願状況（計 0 件）

○取得状況（計 0 件）

〔その他〕

ホームページ等

<http://www.fp.med.kyoto-u.ac.jp/about#re01>

DNA 多型の法医学応用と法数学的解釈

6. 研究組織

(1) 研究分担者

研究分担者氏名：山田 亮

ローマ字氏名：(YAMADA, Ryo)

所属研究機関名：京都大学

部局名：医学研究科

職名：教授

研究者番号（8 桁）：50301106

研究分担者氏名：小谷 泰一

ローマ字氏名：(KOTANI, Hirokazu)

所属研究機関名：京都大学

部局名：医学研究科

職名：准教授

研究者番号（8 桁）：20330582

(2) 研究協力者

研究協力者氏名：眞鍋 翔

ローマ字氏名：(MANABE, Sho)

研究協力者氏名：平井 瑛里子

ローマ字氏名：(HIRAI, Eriko)

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。