

令和 4 年 6 月 7 日現在

機関番号：23903

研究種目：若手研究

研究期間：2019～2021

課題番号：19K19488

研究課題名（和文）X染色体における一塩基多型マイクロハプロタイプの検索と血縁鑑定への応用

研究課題名（英文）Finding single nucleotide polymorphic microhaplotypes on the X chromosome and their application to kinship tests.

研究代表者

福田 真未子（FUKUTA, Mamiko）

名古屋市立大学・医薬学総合研究院（医学）・助教

研究者番号：60832201

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：マイクロハプロタイプと呼ばれる一塩基多型が約200bpの範囲に数個並んだ部分のハプロタイプが、個人識別マーカーとして注目されている。本研究では未だ報告の少ないX染色体に着目し、多型性に富む10座位の新たなマーカーを確立した。また、それらの日本在住者集団のデータ（男性437人、女性236人）収集を行い、集団遺伝学的解析を行った。さらに共有アリル数による血縁関係のシミュレーションにおいても、X染色体の特性を生かした欠損事例等で有用であることが示された。

研究成果の学術的意義や社会的意義

法医学において血縁鑑定に用いるマーカーの新規開拓は、（1）既存の問題点の解決、（2）解析技術の進歩に対応するため、の2点において必要不可欠であり、本研究もその一端を担うものである。マイクロハプロタイプは既存の問題点である「反復配列（STR）のスタターと変異」、 「一塩基多型（SNV）の多型性の低さ」を補う利点を持つことが示された。さらに近年多方面で活用されてきている次世代シーケンサーを利用した法医遺伝学的解析の一方法として、有用である点および新たな問題点を指摘した。

研究成果の概要（英文）：Microhaplotypes, which are haplotypes of several single nucleotide variants arranged in a 200 bp range, have become a focus of attention as genetic markers for individual identification and kinship test. In this study, we investigated the X chromosome, which has not been reported yet, and established 10 new loci with high polymorphism. We also collected Japanese population data (437 males and 236 females) and conducted a population genetic analysis. Furthermore, a simulation study was performed for the kinship test of full-sister by the number of shared alleles using the population data. It showed that the X chromosomal microhaplotypes were useful in the sibship test in females without their parents.

研究分野：法医学

キーワード：DNA多型 マイクロハプロタイプ 血縁鑑定 個人識別 X染色体

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

(1) 既存の方法における問題点

法医学における個人識別や血縁鑑定では、反復配列 (STR) のリピート数や一塩基多型 (SNV) の多型をマーカーとして利用してきた (表1)。しかし、STR は「多型性に富み識別能が高い」というメリットがあるが、「アーチファクトによりスタター (本来より 1 回短い型) が検出されやすい」、「変異率が高く血縁関係に矛盾が生じやすい」というデメリットがある。一方、SNV は変異率は低い、多型性が低い。そのため多数の座位を調べる必要があり、検査のコストがかかるというデメリットがある。

(2) 新しいマーカー概念; マイクロハプロタイプ

次世代シーケンス解析により、同一染色体上の SNVs の各アリルが相同染色体上にあるかを判別し、ハプロタイプを直接決定することが可能になり、マイクロハプロタイプ (MH) という新しい概念が提唱されている (表1)。MH は 3~5 個の SNV が 200bp 以下の範囲に並んだ構造をしており、SNV を組み合わせていることから比較的高い多型性を期待できる。また MH 内の SNV 間の組み換えは無視できるため、ハプロタイプをアリルとして扱うことができる。したがって適切な MH のセットおよび解析法が得られれば、STR と SNV のメリットをあわせ持つマーカーとして活用できると考えられた。

表1. 法医学で用いられる遺伝マーカーの種類と特徴.

マーカーの種類	Short tandem repeats (STR)	Single nucleotide variant (SNV)	Microhaplotype (MH)
形態例	-ATC ATC ATC- -ATC ATC ATC ATC-	-A- -G-	-A-T-C- -T-C-G-
アリル	繰り返し回数 3/4	塩基 A/G	ハプロタイプ ATC/TCG
多型性	高い	低い	やや高い
変異率	高い	低い	低い

2. 研究の目的

MH は近年現れた概念であり、常染色体のデータが蓄積されつつある状況にあるが、X 染色体ではまだ具体的な座位を示しての報告はない。したがって本研究の目的は新規に血縁鑑定用マーカーとして適した MH を検索するものである。また、日本在住者集団のポピュレーションデータを収集することにより、マーカーとしての性質も解析可能であると考えられた。

まず、高い識別能を有する X 染色体 MH を検索し、日本在住者集団におけるハプロタイプ頻度を収集する。法医学において血縁鑑定用のマーカーの新規開拓は、既存の問題点の解決および解析技術の進歩に対応するために必要不可欠であり、本研究もその一端を担うものである。MH は既存の問題点である「STR のスタターと変異」、「SNP の多型性の低さ」を補う利点を持つと考えられた。

次に、MH の法医学多型マーカーとしての性質を明らかにする。標本が適切であるかの指標である「ハーディワインバーク平衡」や、連鎖がマーカー同士の関連に与える影響の指標である「連鎖不平衡」の解析などにより各 MH の性質を検討する。女性データを用いたハーディワインバーク平衡の解析、男性データを用いた連鎖不平衡の解析などを行う。血縁鑑定に有用な MH の座位セッ

トおよび日本在住者集団のポピュレーションデータを公表することで、実務応用も期待できるものと考えられた。

3. 研究の方法

(1) 公開ゲノムデータを用いた座位の選定

本研究は日本での実務応用を想定し、MHの選定は日本在住者集団を対象として行う。MHの定義に沿って、対象箇所は非コード領域で組み換え確率の高いホットスポットを除き、200bp以内に3~5個のSNVを含み、ハプロタイプが3個以上ある箇所とした。公開データの「1000ゲノム」に含まれる105人分の日本人ハプロタイプデータで異型接合度0.5以上、さらにその他のヨーロッパ、アメリカ等でも異型接合度0.4以上の多型性がある箇所を選定した。また検査試薬の都合上、十数か所に絞る必要があったため、3Mb以上の間隔で16組(座位)選択し、対象座位増幅のためのカスタムプライマーを作成した。

(2) ポピュレーションデータの収集

試料は同意を得た日本人集団673人(男性437人、女性236人)の血液から抽出したDNAを用いた。イオン半導体を利用した次世代シーケンサー(Ion GeneStudio S5)はライブラリ作成、増幅産物の定量、テンプレート調整、シーケンシングの順で実施するが、マニュアルおよび機器による実施が選択できるステップがある。本研究ではカスタムプライマーを用いてマニュアルでライブラリ作成を行い、定量PCRで定量後、機器(Ion Chef)によりテンプレート調整し、シーケンサーにかけた。解析には機器専用プラグインを使用し、シーケンス反応が成功しているか否か(総リード数10000以上、On target率30%以上、平均深度500以上)、また各アレルのリード数の算出を行った。また5試料では従来機器であるキャピラリーシーケンサーにより、個別の座位毎にアレルの検証を行い、次世代シーケンサーの結果と一致することを確認した。

(3) 頻度データの解析と評価

得られた集団データから、個人識別や血縁鑑定に有用であるかを評価するための法医学的な指標値を算出した。さらに確率的評価を行う上で確認が必要となるハーディワインバーク平衡からの乖離の有無、型頻度の男女差の有無、MH間の連鎖不平衡の有無を調べた。さらに頻度データを元に、血縁鑑定事例を想定したコンピュータによるシミュレーションを行い、有用性を評価した。

また、法医学実務で重要なアーチファクトについて検討した。次世代シーケンサーは約3%のシーケンシングエラーがあることが知られており、そのようなエラーによる実際には存在しないリードがどの程度検出されるのかについて集計した。

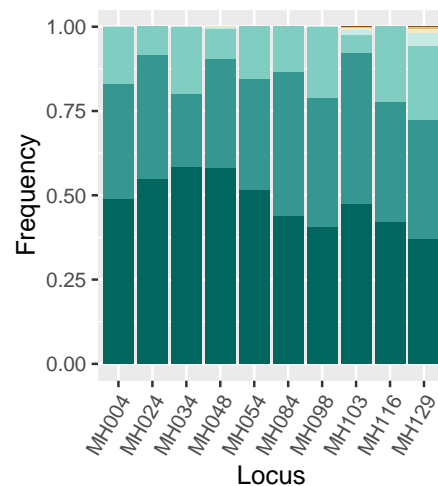


図1. X染色体マイクロハプロタイプ10座位のアレル頻度分布. 下矢印は座位の位置を表す.

4. 研究成果

(1) ポピュレーションデータと各種指標値

対象の16座位のうち10座位で良好なデータが得られた。データ不良の座位の一部ではシーケンシング試薬に含まれるアダプター配列によるシーケンシング反応の障害が疑われた。また、公開されているリファレンスゲノムデータと配列が一致しない座位も存在し、非コード領域におけるリファレンスゲノムデータには、特に比較的長い挿入等の未検出多型が存在することがわかった。さらに、解析ソフトの種類や設定によってはSNVの位置に齟齬が生じる可能性が認められ、解析の

際には条件を共通にすることが重要であることがわかった。

10 座位のアリル頻度分布を図 1 に示す。多型性の指標値である effective number of alleles; Ae の平均値は 2.57, 異型接合度は 0.61 であった (表 2)。ハーディ・ワインバーグ平衡からの乖離の程度である F_{ST} 値は, Bayes 推定による推定値および Hz から算出した概算値の 2 種を算出したが, いずれも大きな乖離はなかった (図 2 左)。また, 隣り合う座位間の連鎖不平衡の程度についても 2 種の指標値を算出したが, 大きな不平衡は存在しなかった (図 2 右)。

表 2. X 染色体 10 座位における多型性の指標値.

座位	MH004	MH024	MH034	MH048	MH054	MH084	MH098	MH104	MH116	MH129
Ae	2.61	2.26	2.34	2.23	2.51	2.55	2.81	2.34	2.83	3.22
Hz	0.617	0.557	0.573	0.551	0.602	0.607	0.644	0.573	0.646	0.689

Ae: effective number of alleles, Hz: Heterozygosity

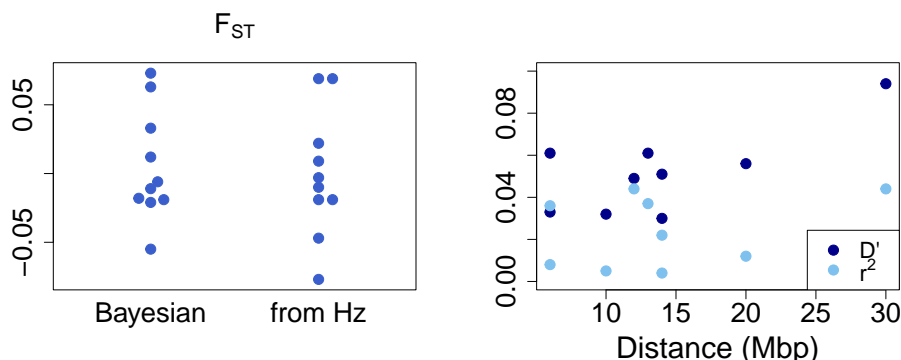


図 2. X 染色体マイクロハプロタイプ 10 座位における指標値. 左: ハーディ・ワインバーグ平衡からの乖離の程度, 右: 連鎖不平衡の程度

(2) 血縁鑑定への応用

研究開始当初は確率的評価を想定していたものの, 次世代シーケンサーの特性上, MH を検査する場合は多数の座位を検査すると考えられ, 確率的な評価は煩雑であること, 共通アリルの個数である IBS 数のみでも十分な識別能を有することがわかったため, IBS 数による評価を行った。姉妹間の同胞鑑定について 10 座位のアリル頻度を用い, 10000 組の全同胞および非血縁のペアを生成しシミュレーションを行ったところ, 約 90% の非血縁ペアで血縁関係が否定された (図 3)。本 X 染色体 MH を常染色体の MH と組み合わせることによって欠損事例の血縁鑑定において有用であると考えられた。

(3) ノイズに関するデータ

男性データを用い, シーケンシングエラーにより生じたノイズリードのうち, 最も頻度が高かったものについてリード頻度を集計したところ, 図 4 のようであった。およそ 2% 以下に収まっている座位が多かったが, 二峰性の分布を示し, やや高い頻度を呈する座位も存在した。STR のスタターのように繰り返し回数に比例して増加するといった単純な法則ではなく, アリルによって傾向が異なることもわかった。このようなノイズの情報には, 既に報告されている常染色体の MH でも十分にデータが蓄積されているとはいえず,

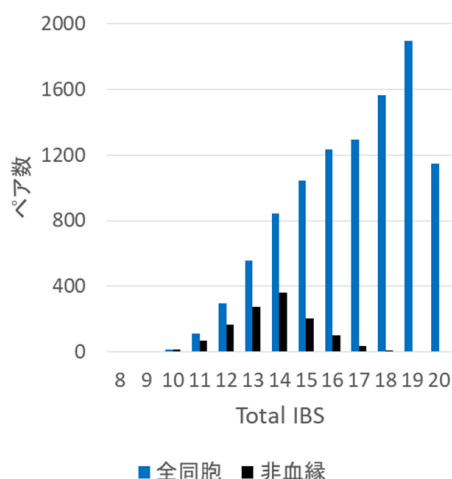


図 3. X 染色体マイクロハプロタイプ 10 座位を用いた全同胞姉妹鑑定における共通アリル (IBS) 数の分布. 10,000 組のシミュレーションによる.

ノイズを生じやすい座位の特徴等についてはさらなる探求が必要であると考えられた。

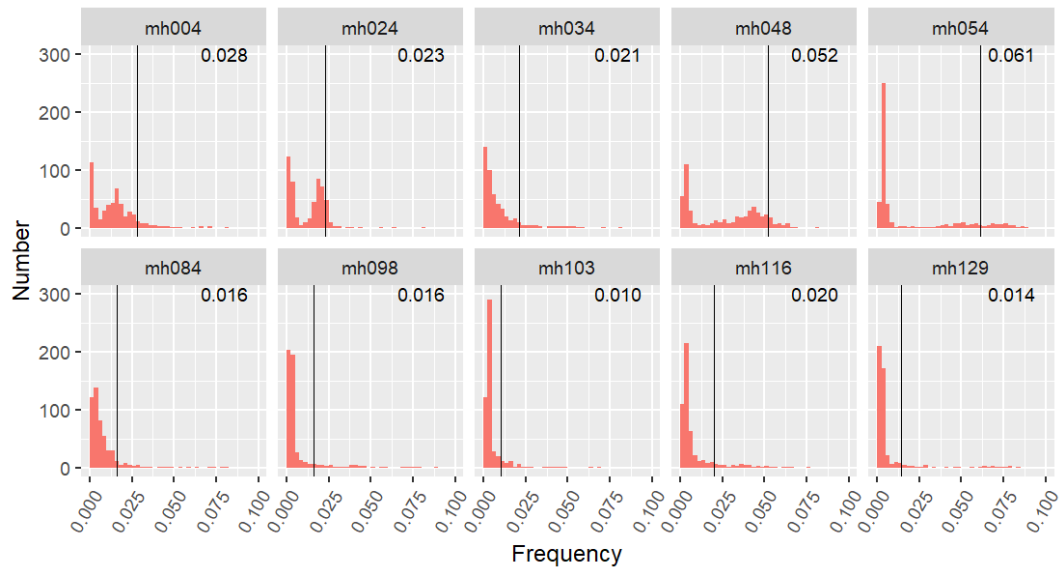


図4. X 染色体マイクロハプロタイプ 10 座位におけるノイズリードの頻度分布. 縦線および右上の数値は 90 パーセンタイル値を表す.

(5) まとめ

以上の結果から、X 染色体 MH は実務応用にも利用可能と考えられる多型性を持ち、特に常染色体では判別困難な欠損事例の血縁鑑定で有用であることが示された。一方で、ノイズリードの性質等についてはアレル毎に特徴がある可能性があることがわかった。既に報告されている座位を含め、アーチファクトの傾向のさらなる情報の蓄積が必要であると考えられた。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 福田真未子, 加藤 秀章, 青木 康博	4. 巻 28
2. 論文標題 ベイズ推定によるポピュレーションデータにおけるfixation indexの推定及びその評価	5. 発行年 2020年
3. 雑誌名 DNA多型	6. 最初と最後の頁 136-139
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 福田真未子, 加藤 秀章, 青木 康博	4. 巻 29
2. 論文標題 STR 配列多型が個人識別における確率的評価に与える影響について	5. 発行年 2021年
3. 雑誌名 DNA多型	6. 最初と最後の頁 87-90
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 福田真未子, 加藤 秀章, 青木 康博	4. 巻 30
2. 論文標題 単一資料を用いたマイクロハプロタイプ解析時に出現するノイズの性質	5. 発行年 2022年
3. 雑誌名 DNA多型	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 福田真未子, 加藤 秀章, 青木 康博
2. 発表標題 STR 配列多型が個人識別における確率的評価に与える影響について
3. 学会等名 日本DNA多型学会第29回学術集会
4. 発表年 2020年

1. 発表者名 福田真未子, 加藤 秀章, 青木 康博
2. 発表標題 単一資料を用いたマイクロハプロタイプ解析時に出現するノイズの性質
3. 学会等名 日本DNA多型学会第30回学術集会
4. 発表年 2021年

1. 発表者名 福田真未子, 大瀧 純, 中村昌美, 菅野さな枝, 加藤 秀章, 青木 康博
2. 発表標題 X染色体マイクロハプロタイプ10座位に関するポピュレーションデータとその評価
3. 学会等名 第105次日本法医学会学術全国集会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------