

令和 4 年 4 月 28 日現在

機関番号：12601

研究種目：挑戦的研究（萌芽）

研究期間：2018～2021

課題番号：18K19426

研究課題名（和文）マイクロサテライトを用いた人類集団史の推定

研究課題名（英文）Comprehensive analysis of microsatellites polymorphism in the human population

研究代表者

藤本 明洋（Fujimoto, Akihiro）

東京大学・大学院医学系研究科（医学部）・教授

研究者番号：30525853

交付決定額（研究期間全体）：（直接経費） 4,500,000円

研究成果の概要（和文）：本研究では、SNPに代わり、突然変異率が高いマイクロサテライトを用いることで、近縁な集団間の遺伝的関係性や最近の集団史の情報を得ることを目的とした。藤本が作成した手法（Fujimoto et al. Genome Research (2020)）を用いて、900万個以上のマイクロサテライトを対象に多型的なマイクロサテライトを解析した。マイクロサテライトを用いて主成分分析（PCA）を行ったところ、先行研究（SGDP Nature (2016)）と比べ、本研究の方が、SNPとの類似性が高く、マイクロサテライト解析が先行研究よりも精度が高いことが示唆された。

研究成果の学術的意義や社会的意義

近年のゲノム解析技術の進展により、DNA多型解析アレイや全ゲノムシーケンスを用いて大量のSNP情報を得ることができる。この情報を用いて、人類の集団史を推定することができる。例えば、集団間の遺伝的関係性の推定や、過去の集団の大きさの変遷が推定されている。しかし、SNP情報を用いた研究では、極めて近い集団の関係性や最近の集団サイズの変遷を正確に推定することが難しい。本研究では、突然変異率が高いマイクロサテライトを用いることで、近縁な集団間の遺伝的関係性や最近の集団史の情報を得ることを目的とした。本研究は、マイクロサテライトを用いた集団解析のための方法論を提供する。

研究成果の概要（英文）：Microsatellites have been widely used as genetic markers in population genetics and genealogy studies due to their high mutation rate and rich patterns of repeat length among individuals. However, few studies have been analyzed the genome-wide pattern of microsatellites due to methodological difficulties. To reveal the landscape of the polymorphism in microsatellites at the whole genome level, we analyzed 9 million microsatellites in approximately 4,000 whole-genome sequencing (WGS) data of four public datasets (HGDP, SGDP, ICGC, KPGP) across the various ethnic groups. We analyzed the genetic differentiation among human populations. The principal component analysis (PCA) of microsatellites showed the efficiency of microsatellites for genetic studies. Our analysis provides a comprehensive picture of microsatellite polymorphisms along with the detection method, as well as the efficiency in population genetic studies.

研究分野：人類遺伝学

キーワード：マイクロサテライト 人類の集団史

1. 研究開始当初の背景

人類の集団史の推定は、人類学的な重要性だけでなく、疾患関連遺伝子の探索の背景知識としても重要な課題である。近年のゲノム解析技術の進展により、DNA 多型解析アレイや全ゲノムシーケンセスを用いて大量の SNP 情報を得ることができる。この情報を用いて、人類の集団史を推定することができる。例えば、集団間の遺伝的関係性の推定や、過去の集団の大きさの変遷が推定されている。

しかし、SNP 情報を用いた研究では、極めて近い集団の関係性（例えば、本州の中の地域間の関係）や最近（1000 年以降）の集団サイズの変遷は、正確に推定することが難しい。この理由の一つは、マーカーとして SNP を用いているためであると考えられる。SNP の突然変異率低く、 1.2×10^{-8} /世代程度であると推定されている。そのため、安定性が高いものの、最近の集団間の遺伝的变化を捉えるマーカーとしての能力は限られていると考えられる。

2. 研究の目的

本研究では、SNP に代わり、突然変異率が高いマイクロサテライトを用いることで、近縁な集団間の遺伝的関係性や最近の集団史の情報を得ることを目的とした。

3. 研究の方法

申請者らが開発したマイクロサテライト多型検出法を用いて、公開されている人類集団のシーケンセスデータを解析した。

4. 研究成果

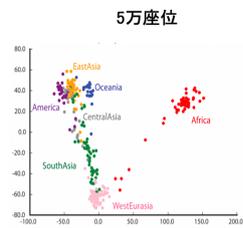
ヒトゲノム内には様々な繰り返し配列が存在しているが、どの領域をマイクロサテライトとするかは、定義する手法により異なる。また、我々は、次世代シーケンサーのリード情報からマイクロサテライトの遺伝的多様性を検出することを目的としているため、次世代シーケンサーの解析に適したマイクロサテライトを選出する必要がある。そこで、我々は RepeatMasker, Tandem Repeat Finder, MISA の 3 種類のソフトウェアを用いて 15,737,726 個のマイクロサテライトを定義した。その後、次世代シーケンサーでの解析のため、前後の領域の配列がユニークであり、近傍に他のマイクロサテライトがなく、長さが 80bp 以下のマイクロサテライトを選出した。この結果、8,817,054 個のマイクロサテライトが選出された。

次にマイクロサテライトの多型を検出する方法を開発した。マイクロサテライトには A の連続、AC の連続などさまざまな塩基のパターンが存在している。それぞれのマイクロサテライトのパターンごとにシーケンサーエラー率を推定した。エラー率の推定には、ヘミ接合でありヘテロ接合が存在しない男性の X 染色体のデータを用いた。その結果、A/T の割合が高いマイクロサテライトや繰り返し単位が短いマイクロサテライトはエラー率が高いこと、欠失のエラーは挿入のエラー率よりも高いことが分かった。推定したエラー率を考慮する確率モデルを用いて、エラーと真の変異を区別する方法を開発した。解析の精度は、擬似データおよび実験を用いて評価した。

この結果に基づき、藤本が作成した手法 (Fujimoto et al. Genome Research (2020)) を用いて、900 万個以上のマイクロサテライトを対象に多型的なマイクロサテライトを解析した。韓国人データ (KPGP)、世界各地の集団のデータ (SGDP) のデータを解析した。KPGP および SGDP のデータから、約 28 万と 42 万個の多型的な ($MAF \geq 3\%$) マイクロサテライトが発見された。また、長いマイクロサテライトほど多型性が高いこと、コード領域には繰り返し単位が 3 の倍数のマイクロサテライトが多いこと、AT 含有量が高いマイクロサテライトほど多型性が高いことなどが明らかになった。

マイクロサテライトを用いて SGDP のデータに対して主成分分析 (PCA) を行った。先行研究 (SGDP Nature (2016)) の PCA と比べ、本研究の方が、SNP との類似性が高く、今回のマイクロサテライト解析が先行研究よりも精度が高いことが示唆された (図 1)。現在は、ICGC (国際がんゲノムコンソーシアム) のデータの解析を行っている。

先行研究 (SGDP Nature(2016))



本研究

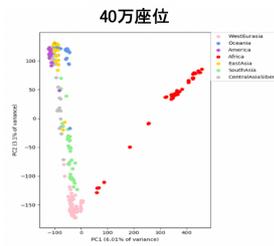


図1 SGDPのデータを用いたPCAの結果。先行研究と比較すると、本研究の方が、解像度が高いことが示唆される

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 2件/うちオープンアクセス 1件）

1. 著者名 Fujita M, Yamaguchi R, Hasegawa T, Shimadad S, Arihiro K, Hayashi S, Maejima K, Nakano K, Fujimoto A, Ono A, Aikata H, Ueno M, Hayami S, Tanaka H, Miyano S, Yamaue H, Chayama K, Kakimi K, Tanaka S, Imoto S, Nakagawa H	4. 巻 53
2. 論文標題 Classification of primary liver cancer with immunosuppression mechanisms and correlation with genomic alterations.	5. 発行年 2020年
3. 雑誌名 EBioMedicine	6. 最初と最後の頁 102659
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ebiom.2020.102659	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 該当する
1. 著者名 Fujimoto A*, Fujita M, Hasegawa T, Wong JH, Maejima K, Oku-Sasaki A, Nakano K, Shiraishi Y, Miyano S, Yamamoto G, Akagi K, Imoto S, and Nakagawa H	4. 巻 30
2. 論文標題 Comprehensive Analysis of Indels in Whole-genome Microsatellite Regions and Microsatellite Instability across 21 Cancer Types	5. 発行年 2020年
3. 雑誌名 Genome Research	6. 最初と最後の頁 334-34
掲載論文のDOI (デジタルオブジェクト識別子) 10.1101/gr.255026.119	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium.	4. 巻 578
2. 論文標題 Pan-cancer analysis of whole genomes.	5. 発行年 2020年
3. 雑誌名 Nature	6. 最初と最後の頁 82-93
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41586-020-1969-6	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 Mizuno K, Akamatsu S, Sumiyoshi T, Wong J1. Mizuno K, Akamatsu S, Sumiyoshi T, Wong JH, Fujita M, Maejima K, Nakano K, Ono A, Aikata H, Ueno M, Hayami S, Yamaue H, Chayama K, Inoue T, Ogawa O, Nakagawa H and Fujimoto A*	4. 巻 9
2. 論文標題 eVIDENCE: a practical variant filtering for low-frequency variants detection in cell-free DNA.	5. 発行年 2019年
3. 雑誌名 Sci Rep	6. 最初と最後の頁 15017
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-019-51459-4	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 Wong JH, Shigemizu D, Yoshii Y, Akiyama S, Tanaka A, Nakagawa H, Narumiya S, Fujimoto A	4. 巻 11
2. 論文標題 Identification of intermediate-sized deletions and inference of their impact on gene expression in a human population.	5. 発行年 2019年
3. 雑誌名 Genome Medicine	6. 最初と最後の頁 44
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s13073-019-0656-4	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0件

〔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
---------	---------