

令和 5 年 6 月 16 日現在

機関番号：82626

研究種目：若手研究

研究期間：2019～2022

課題番号：19K16246

研究課題名（和文）迅速なDNA多型解析システムの開発

研究課題名（英文）Development of a rapid DNA typing system

研究代表者

石谷 孔司（Ishiya, Koji）

国立研究開発法人産業技術総合研究所・生命工学領域・研究員

研究者番号：40826062

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）：本研究課題では、次世代型のDNAシーケンサーに対応した迅速なDNA多型解析システムの構築に向けて、ヒトのミトコンドリアゲノム、Y染色体、性別推定等に重要なマーカーや指標の開発に加え、微生物をはじめとするヒト以外の生物種を対象としたDNA多型解析技術も開発した。本課題で提案するY染色体ハプログループの推定法やヒト性別判定法については、数十塩基程の短い塩基配列をターゲットとすることで従来法に比べ迅速な判定が可能でありながら、シーケンスエラーやゲノムカバレッジに影響されにくいことから、今後のDNA多型解析の基盤技術として利用が期待される。

研究成果の学術的意義や社会的意義

PCR法を前提とした従来法では限定的なDNA多型領域を用いて性別・品種・親子等の判定が行われていたが、この方法は実験手法を含めて判定までに非常に時間がかかる方法であった。本課題における提案手法では次世代型シーケンサーの特性である網羅性を活かした迅速な判定が可能であるため、今後フィールドワークでの簡易的なDNA検査や生物種の判定、災害時における迅速な身元判定、医療現場における迅速なDNA検査の基盤となる技術として期待できる。

研究成果の概要（英文）：Toward the development of a rapid DNA polymorphism analysis system compatible with next-generation DNA sequencers, important markers and indicators were developed for the human mitochondrial genome, Y chromosome, and gender estimation. In addition, DNA polymorphism analysis technologies were developed for species other than humans, including microorganisms. The Y chromosome haplogroup estimation and human sex identification methods proposed in this project are characterized by the fact that they are faster than conventional methods by targeting short sequences and are robust to sequencing errors and variations in genome coverage.

研究分野：バイオインフォマティクス

キーワード：DNA多型 バイオインフォマティクス 次世代シーケンサー 性別判定

様式 C-19、F-19-1、Z-19（共通）

1. 研究開始当初の背景

研究開始当初は、第一世代のキャピラリー型 DNA シークエンサーから第二世代、第三世代、第四世代と次世代型の DNA シーケンサー (Next Generation Sequencer; NGS) へと置き換わりつつあった。この背景を受けてヒトや動植物を対象とした DNA 鑑定においても NGS の特性や性能を活かした新たな DNA 多型解析技術が必要とされていた。従来までの DNA 多型解析の主な対象となっていたのは、STR (Short Tandem Repeat; STR) と呼ばれる短い繰り返し配列や一塩基多型 (Single Nucleotide Repeat; SNP) 等の限定的な領域であった。こうした DNA 多型は、PCR 法をはじめとする DNA 増幅法によって特定領域のみを増幅し分析するため、NGS の特性の一つである網羅性あるいはナノポア DNA シーケンサーに代表されるようなリアルタイム性を十分に活かされていなかった。

2. 研究の目的

上記のような背景を踏まえ、本研究課題では次世代シーケンサーの特性を活かした迅速な DNA 多型解析システムの構築を行うべく、新たな DNA 多型のマーカー開発やそれに伴う情報解析技術の開発を目的とした。

3. 研究の方法

本研究課題では主に下記の(1)～(4)の項目について研究開発を実施した。

(1) ヒトミトコンドリアゲノム多型の補完技術

ヒトの系統推定に重要なミトコンドリアゲノムを対象とした DNA 多型解析技術として、DNA 多型を高速かつ高精度に補完するための技術開発を行った。ヒトのミトコンドリアゲノムは母系の系統関係を調べる上で有益な遺伝情報を有しており、世界各地のヒト集団において約 5,000 以上の系統 (ハプログループ) が知られている。しかしながら、法医学試料や考古試料等の劣化試料や微量試料においては、試料の保存状態によってはミトコンドリアゲノムの部分的な配列しか得られず試料の比較解析が十分に行えないケースがあった。そこで、本研究課題では、ミトコンドリアゲノムにおける多型情報を元に機械学習モデルを構築し、欠損したゲノム領域を高精度に補完する手法として MitoIMP を開発した。

(2) ヒト Y 染色体ハプログループの高速推定技術

ヒトの Y 染色体は、父系の系統関係を調べる上で重要な領域であり、世界各地の様々な地域集団で調べられている。NGS データからヒト Y 染色体のハプログループを推定する通常の解析では、ゲノムマッパーを用いて次世代シーケンサーから得られたリード配列をヒトゲノム参照配列にマッピングし、SNP をコールする。そして、この呼び出された SNP のジェノタイプング結果に基づいて Y ハプログループを割り当てる。一方、本研究のアプローチでは、各 Y 染色体ハプログループに特異的な K-mer 配列からハプログループの推定を直接行うことで、従来法に比べて迅速なヒト Y 染色体ハプログループの推定を可能とした。

(3) 微生物に対するゲノム多型解析技術

法医学試料や考古試料等にはヒト以外の生物種 (主に土壌や環境中の微生物) 由来の DNA (メタゲノム) が大量に含まれる。こうしたメタゲノムは個人間 (サンプル間) 差が大きいことから、個人識別等での応用が期待されている。本研究課題では、これらの微生物に対するゲノム多型解析技術も併せて開発した。これらは、試料周囲の環境変化や風化の影響を受けた試料の調査、試料に付着している非ターゲット生物の DNA 組成、コンタミネーションのプロファイリング、病原性微生物の有無に対する調査等に応用可能である。

(4) 迅速なヒト性別判定技術

ヒトの X/Y 染色体上にあるアメロゲニン遺伝子は、その配列長に性染色体 (X/Y) 上で違いがあることから、この長さや数の違いにより性別が推定できる。この方法は、試料中の内在性 DNA が PCR 増幅でき、DNA の断片化が生じていない場合は適用できるが、長い期間で風化した劣化試料や考古試料に対する適用性は低い。最近では次世代シーケンサーデータのリード配列 (読み取られた塩基配列) の X/Y 染色体へのマッピングされたリード比 (R_y) に基づく性別判定法も提案されているが、ヒト全ゲノムへのマッピング工程が前提であり、解析作業に時間を要し、さらに低カバレッジのサンプルでは誤差が大きいとされている。本研究課題では、この R_y を K-mer 配列に拡張した KR_y という新たな性別判定の指標を開発し、リード配列から直接推定できる迅速な性別推定アルゴリズムを開発した。なお、この KR_y は、女性であれば 0、男性であれば 0.5 に限りなく近い値を示すよう設計されている。

4. 研究成果

(1) ヒトミトコンドリアゲノム多型の補完技術

本研究課題で開発した MitoIMP を複数のミトコンドリアゲノム系統で検証したところ、全体の 65%以上が欠損した場合であっても、欠損領域を 99%以上の精度で補完できることが分かった(図 1)。この手法により、NGS 等で得られた複数個体のヒトミトコンドリアゲノムの部分配列を互いに比較解析することができるようになった。なお、開発した欠損補完アルゴリズム MitoIMP を実装したプログラムは、Python 言語でオープンソース化されており、Github 上で広く公開されている。また、MitoIMP は距離計算時の速度改良として、並列処理技術と実行時コンパイラ機能等を併用し、通常の Python コードプログラムに比べて高速な動作を実現している。MitoIMP のアルゴリズムの概要とその成果は、査読付きのオープンアクセス国際誌で公開されている。

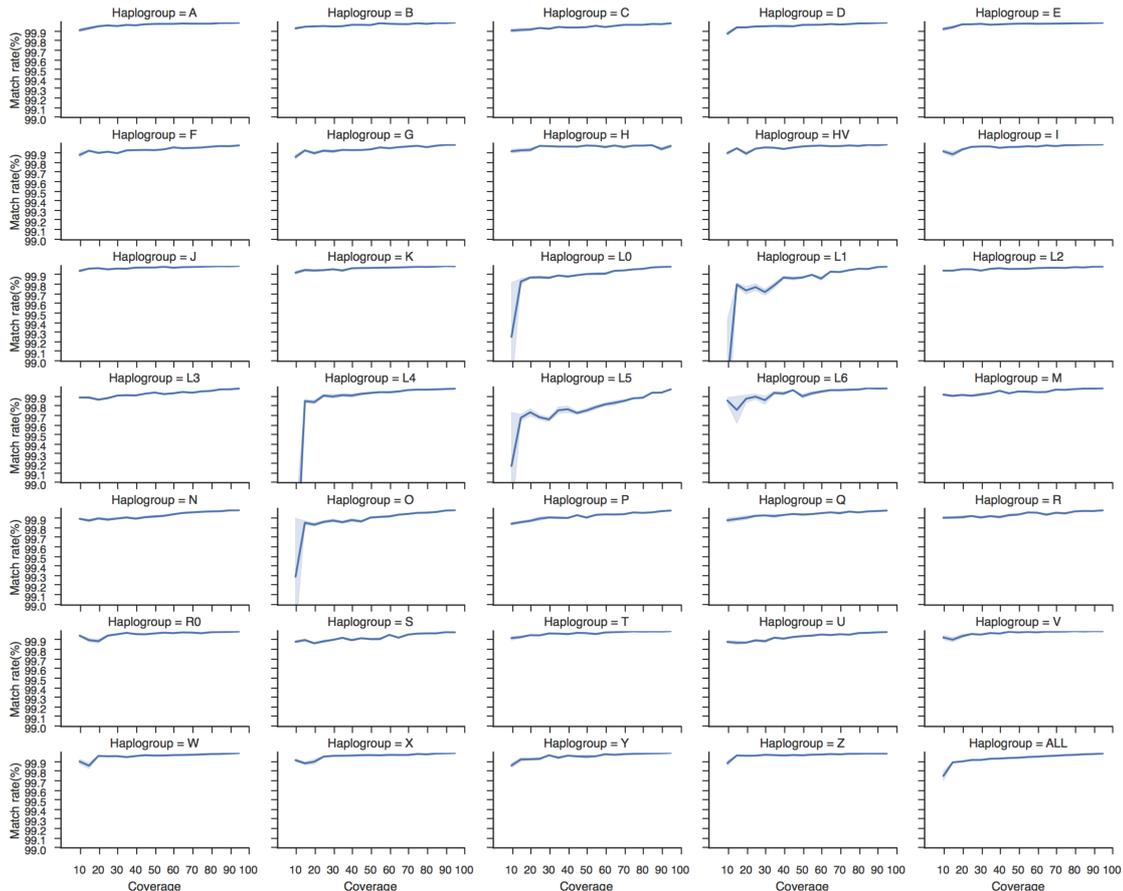


図 1. 様々なヒトミトコンドリアゲノム系統における欠損補完精度

(2) ヒト Y 染色体ハプログループの高速推定技術

本研究課題における提案手法では、Yleaf (Ralf *et al.*, 2018) による最新手法と同等のハプログループを推定することができ、さらに 10 倍以上速く Y 染色体の SNP コールと Y ハプログループの推定が可能であることが示された。以上の成果は、欧州進化学会主催の国際会議にて成果発表を行った。

(3) 微生物に対するゲノム多型解析技術

本課題では、ヒト以外の生物種(微生物など)のゲノムを高速に比較分析可能な技術(解析ツール)も併せて開発を行った。この技術では、ゲノム間の違い(類似性)を二次元空間上に距離として表現することで、視覚的かつ高速にゲノム比較や多型の有無を確認することが出来る。開発されたツールは Github 上で公開されている。

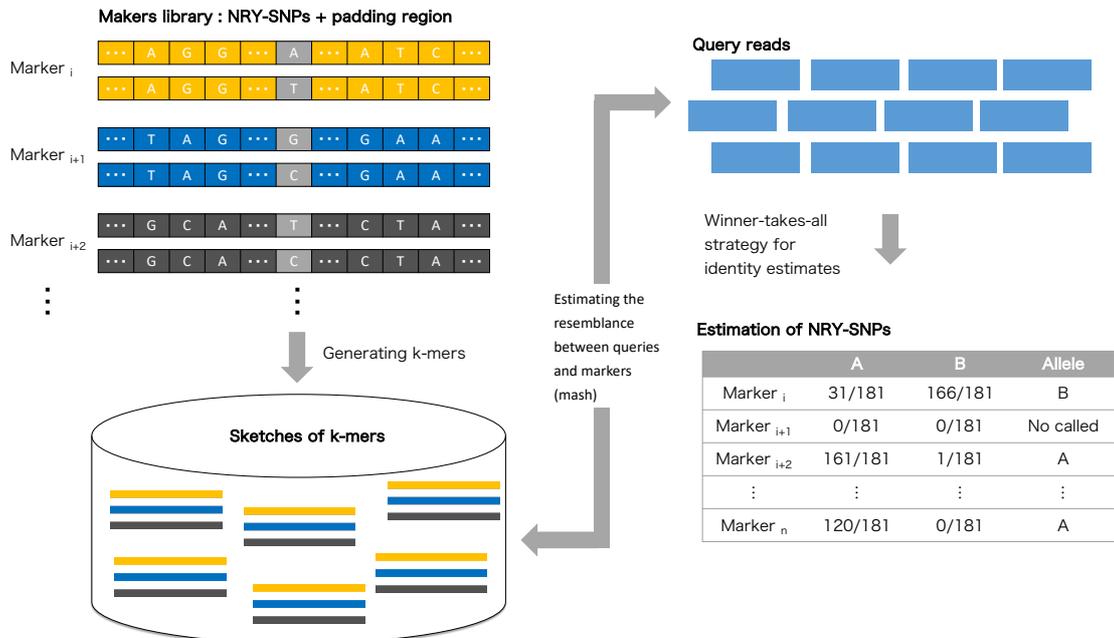


図 2. Y 染色体ハプログループの推定方法（提案手法）の模式図

(4) 迅速なヒト性別判定技術

本提案手法を最も広く利用されている illumina の次世代シーケンスデータ、さらにポータブル型 DNA シークエンサー MinION から得られたデータにそれぞれ適用し、その有効性を検証した。illumina シーケンサーを想定した 5X のヒト全ゲノムデータ（シミュレーションデータ）を用いて検証したところ、本提案手法は、NGS データを用いた性別判定法として最も利用されている Skoglund *et al.*, (2013) の方法に対して、約 1/40 の時間で性別判定が可能なが示された。また、従来法では判定が難しかった 0.001X の非常にゲノムカバレッジの低いデータの場合においても Ry の値に大きな変動はなく、正確な性別判定が可能であった。実データを用いた検証では、性別が既知の 1000 人ゲノムプロジェクトから JPT（日本人）の全ゲノムデータを用いて本提案手法を検証した結果、全てのサンプルで正確な性別判定が可能であった。また、ポータブル型 DNA シークエンサーである MinION から得られたヒトゲノムデータに対しても本手法を適用した結果、KRy に大きな誤差はなく、正確な性別判定が可能であった。以上より本提案手法は、従来法に比べても非常に迅速な判定が可能であり、かつ、カバレッジの影響を受けにくい方法であることが示された。以上の成果は日本 DNA 多型学会にて発表を行い、若手研究賞を受賞した。

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 7件/うち国際共著 3件/うちオープンアクセス 5件）

1. 著者名 石谷孔司、油谷幸代	4. 巻 60(2)
2. 論文標題 バイオインフォマティクスによる微生物群集解析の新展開	5. 発行年 2022年
3. 雑誌名 化学と生物	6. 最初と最後の頁 52～54
掲載論文のDOI（デジタルオブジェクト識別子） 10.1271/kagakutoseibutsu.60.52	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 Ishiya Koji, Nakashima Nobutaka	4. 巻 44
2. 論文標題 Comparative Genome Analysis Reveals Accumulation of Single-Nucleotide Repeats in Pathogenic Escherichia Lineages	5. 発行年 2022年
3. 雑誌名 Current Issues in Molecular Biology	6. 最初と最後の頁 498～504
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/cimb44020034	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 -
1. 著者名 Ishiya Koji, Aburatani Sachiyo	4. 巻 19
2. 論文標題 Multivariate statistical monitoring system for microbial population dynamics	5. 発行年 2021年
3. 雑誌名 Physical Biology	6. 最初と最後の頁 016003～016003
掲載論文のDOI（デジタルオブジェクト識別子） 10.1088/1478-3975/ac3ad6	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 Iwasaki Risa L., Ishiya Koji, Kanzawa-Kiriyama Hideaki, Kawai Yosuke, Gojobori Jun, Satta Yoko	4. 巻 11
2. 論文標題 Evolutionary History of the Risk of SNPs for Diffuse-Type Gastric Cancer in the Japanese Population	5. 発行年 2020年
3. 雑誌名 Genes	6. 最初と最後の頁 775～775
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/genes11070775	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 Mizuno Fuzuki, Ishiya Koji, Matsushita Masami, Matsushita Takayuki, Hampson Katherine, Hayashi Michiko, Tokanai Fuyuki, Kurosaki Kunihiko, Ueda Shintaroh	4. 巻 10
2. 論文標題 A biomolecular anthropological investigation of William Adams, the first SAMURAI from England	5. 発行年 2020年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 21651
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-020-78723-2	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 Koji Ishiya, Fuzuki Mizuno, Li Wang, Shintaroh Ueda	4. 巻 13
2. 論文標題 MitoIMP: A Computational Framework for Imputation of Missing Data in Low-Coverage Human Mitochondrial Genome	5. 発行年 2019年
3. 雑誌名 Bioinformatics and Biology Insights	6. 最初と最後の頁 1~9
掲載論文のDOI (デジタルオブジェクト識別子) 10.1177/1177932219873884	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 Koji Ishiya, Shintaroh Ueda	4. 巻 1391
2. 論文標題 Novel approach for accurate detection of contaminating human mitochondrial DNA in next-generation sequencing data	5. 発行年 2019年
3. 雑誌名 Journal of Physics: Conference Series	6. 最初と最後の頁 012045 ~ 012045
掲載論文のDOI (デジタルオブジェクト識別子) 10.1088/1742-6596/1391/1/012045	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 -

〔学会発表〕 計5件 (うち招待講演 0件 / うち国際学会 3件)

1. 発表者名 Koji Ishiya
2. 発表標題 Time-series Change Detection Approach for Microbial Population Dynamics
3. 学会等名 11th International Conference on Bioscience, Biochemistry and Bioinformatics (国際学会)
4. 発表年 2021年

1. 発表者名 Koji Ishiya
2. 発表標題 Detection of age-related methylation sites for different cell resources
3. 学会等名 EurasianBioChem2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Koji Ishiya
2. 発表標題 Detection of phylogenetically-informative SNPs in human Y-chromosome from next-generation sequencing data
3. 学会等名 The 2019 Congress of the European Society for Evolutionary Biology (国際学会)
4. 発表年 2019年

1. 発表者名 石谷 孔司、水野 文月、熊谷 真彦、五條堀 淳、林 美千子、松下 真実、松下 孝幸、谷口 康浩、近藤 修、黒崎 久仁彦、王 瀝、植田 信太郎
2. 発表標題 居家以岩陰遺跡（縄文早期）及び土井ヶ浜遺跡（弥生中期）出土人骨の核ゲノム解析（予報）
3. 学会等名 第73回日本人類学会大会
4. 発表年 2019年

1. 発表者名 熊谷真彦、水野文月、石谷孔司、五條堀淳、林美千子、松下真実、松下孝幸、谷口康浩、近藤修、黒崎久仁彦、王瀝、植田信太郎
2. 発表標題 居家以岩陰遺跡（縄文早期）及び土井ヶ浜遺跡（弥生中期）出土人骨のエピゲノム解析の試み
3. 学会等名 第73回日本人類学会大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<https://github.com/omics-tools/mitoimp>
<https://github.com/omics-tools/genome-mds>

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
---------	---------