

令和 5 年 6 月 1 日現在

機関番号：13901

研究種目：基盤研究(B) (一般)

研究期間：2017～2021

課題番号：17H04148

研究課題名(和文)ゲノムワイドな反復配列を利用した日本人と日本近隣ヒト集団との識別推定法の確立

研究課題名(英文) Establishment to differentiate an estimating method using genome-wide repeat sequences between Japanese and human populations near Japan

研究代表者

山本 敏充 (Yamamoto, Toshimichi)

名古屋大学・医学系研究科・准教授

研究者番号：50260592

交付決定額(研究期間全体)：(直接経費) 12,400,000円

研究成果の概要(和文)：分散型のAlu配列では、実用的な新規Alu探索法は確立できたが、効率よく断片化することが難しいことが判明した。また、日本人約400人及び韓国人約90人の全ゲノム塩基配列データから得られたAlu挿入箇所のデータにおいて、両集団間で異なり、各集団で20%以上の頻度を持つ挿入箇所を選び評価したが、新しいAlu-Yタイプがあまり多くないことが分かった。一方、縦列型のSTRsでは、261座位のSTRsの遺伝型データを利用して、個人間及び集団間で識別するために必要な座位数は、常染色体の短腕・長腕に各1座位ずつ、個人間でも両集団間でも識別力のある座位を選択すれば、統計学的に識別可能である結果が得られた。

研究成果の学術的意義や社会的意義

反復配列を利用した近縁ヒト集団の遺伝的区別(生物地理学的祖先系の推定)を目的として行った。その結果、日本人と韓国人をAlu配列により区別することは難しいことが判明したことの学術的意義は高い。一方、日本人と漢中国人における261座位のSTRsの遺伝型データの統計・集団遺伝学的解析では、ほぼ常染色体の短腕・長腕に各1座位に相当する計44座位で、個人及び集団間でも識別力のあるSTRs座位により、その識別は可能であることが示唆されたことの学術的意義は高い。また、そのような約45座位による、2つの判定システムを構築すれば、その祖先推定の可能性が得られたことの社会的意義は非常に大きいと考えられる。

研究成果の概要(英文)：Regarding Alu elements, a novel method to apply to search unknown them in human genome was established. However, it was proved difficult to make genome DNA appropriate size of fragmentation. Therefore, using the data of the Alu element positions obtained from whole genome sequence data in about 400 Japanese and 90 Korean, those with different between both populations and with more than 20% of the frequencies in each population were extracted and evaluated. As a result, it was suggested that there were not so much Alu-Y subtypes as more descendant. Alternatively, regarding STRs, we performed some population genetical analyses using the genotype data for 261 autosomal STRs from each 32 Japanese and Han-Chinese. Then, it was suggested it would be possible to differentiate statistically between both populations if we could select each a locus, which has high personal discriminating power and power of the differentiation between both populations. on each arm of each autosomal chromosome.

研究分野：法医学

キーワード：反復配列 Alu配列 Short tandem repeats 日本人 韓族 漢中国人 識別法

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

(1) 人類学の分野では、100 万以上のゲノムワイドな SNPs や全ゲノム配列解析、また、ミトコンドリア DNA の全塩基配列解析が、次世代シーケンサ (NGS) を利用して、現ヒト集団や考古学的な試料からも、比較的容易に、また、費用も年々安価になり、数万円で解析できるようになってきた。一方、分子人類学を応用してきた法医学分野では、個人識別や血縁関係を調べる解析方法は、ほぼ確立され、現在この分野で問題となっている諸問題の多くは、結果の解釈や統計学的な評価である。しかし、これらは全て対照試料がある場合の比較解析である。

(2) しかし、その一方で、ヒトの表現型などをゲノム情報から検出し、個人の遺伝的形質の一部の情報を得るような (FDP: Forensic DNA Phenotyping) 試みが、ヨーロッパを中心としてなされ、マルチプレックスシステムにより虹彩・皮膚・毛髪の色などを特定できるような SNPs の型判定が行われ、ある程度の成果を得てきている。また、最近では身長あるいは顔貌の形質の一部に関連する遺伝子もゲノムワイド関連解析研究 (GWAS) により、特定されてきている。しかし、このような FDP は、特に歴史的にアフリカ系の人種差別問題が根強いアメリカ社会では倫理的な問題も提起されている。ただ、そのようなアメリカでも、200 程度の SNPs を型判定するものの個々のマーカーの情報は詳述せず、全体的な確率論として、アフリカ系・ヨーロッパ系・東アジア系・アメリカ先住民系の混合比、いわゆる生物地理学的祖先を調べるようなビジネスはあるが、実際に、犯罪などの法科学的事例に応用されることはないといって過言でない。

(3) 東アジアでは、日本の近縁諸国の韓国や中国の人々とは、形質的には非常によく似ているが、ゲノムワイドの SNP 解析や全ゲノム解析により、これらのヒト集団とを遺伝的に区別することは可能になってきている。しかしながら、法科学的に応用するために、低分子化した微量な DNA 試料から、裁判を維持できるような精度と、確率論的考察ができる方法は現在のところ見当たらない。研究代表者は、今までに、SNPs よりも突然変異率が高い約 100 の縦列型の反復配列で、1 リピート単位が 4 塩基であるマイクロサテライト (STRs: short tandem repeats) を用いて、日本人と中国人との間での遺伝的な確率論的区別や、より遺伝的に近縁である韓国人 (韓族) との区別について、第 69 回日本人類学会大会 (平成 27 年 10 月開催) 等で、確率論的に区別できる可能性について発表してきた。また、日本人と中国人における約 250 マーカーの STRs の型判定結果もデータとして保持している。さらに、平成 25 年度～27 年度の科学研究費補助金 (挑戦的萌芽研究) を利用して、より突然変異率が高く、レトロトランスポゾンの一つである SINEs (short interspersed nuclear elements) の多くを占める分散型反復配列である Alu 配列の新規部位探索法であるレトロトランスポゾン・ディスプレイ法をヒトにおいてほぼ確立した。そこで、本研究では、遺伝的には中立で、遺伝子と直接連鎖しないと考えられる、これらの縦列反復配列である STRs のデータ及び分散型反復配列である Alu 配列を、日本人と遺伝的に近縁なヒト集団の確率論的区別に利用してはどうかという着想に至った。

2. 研究の目的

(1) ヒトゲノム上の反復配列のうち、現在保有している縦列反復配列である 4 塩基リピートの約 250 の STRs から、日本人と遺伝的に近縁なヒト集団である韓国人 (韓族) あるいは中国人 (漢族) とを区別可能な座位を可能な限り少数選択し、それらを効率よく、また低分子化した DNA 試料から正確に型判定できるマルチプレックスシステムを構築する。

(2) 分散型反復配列のうち、特にヒトゲノム中に 100 万か所以上あるといわれているが、個人間やヒト集団間でゲノム上での分布が非常に異なるため、未だ特定できない Alu 配列に着目し、個々の配列の分布を特定するのではなく、日本人と日本周辺のヒト集団 (特に韓国人) との間において、相違を見せる Alu 配列の新規選択を行い、これを指標として各集団間を確率論的に区別できる部位を選択し、選択された部位を効率よく、また低分子化した DNA 試料から正確に型判定できるマルチプレックスシステムを構築する。

(3) これら 2 つの型判定法を利用して、各集団間の確率論的推定法を確立しようとする事、できれば、将来的に、白骨などの実際の法科学的資料に応用し、その有用性を検討することを目的とする。

3. 研究の方法

(1) 今までの研究で蓄積されていた、5 地域 (秋田、名古屋、大分、長崎及び沖縄) の日本人 160 人、並びに 2 地域 (ソウル及び瀋陽) の韓族 64 人から得られた縦列反復配列である STR (short tandem repeat) の 105 座位のデータを利用して、統計学的に判別可能かどうか検討した。

(2) 今までの研究で蓄積されていた、日本人と漢中国人(各 32 名)における 261 座位の STRs の遺伝型データの正確性について再確認した後、それらの精度の高いデータを利用して、以下のよう統計遺伝学的解析及び集団遺伝学的解析を行った。

- GenAIEx6.503 ソフトウェアを用いた両集団の各座位におけるアレル頻度の算出
- GenAIEx6.503 ソフトウェア(2 検定)及び Arlequin 3.5.2.2 ソフトウェア(Exact test)を用いた両集団の各座位における HWE (Hardy-Weinberg equilibrium) 検定
- STRAF 1.0.5 ソフトウェアを用いた両集団の各座位におけるヘテロ接合度(観察値)などの法科学的統計値の算出
- Arlequin 3.5.2.2 ソフトウェア(Exact test)を用いた各座位における両集団間のアレル頻度分布の有意差検定
- STRAF 1.0.5 ソフトウェアによる主成分分析 (PCA) 及び Genetix 4.05 ソフトウェアによる因子対応分析 (FCA)
- 尤度比による両集団の識別の試み

(3) 国立遺伝学研究所のスーパーコンピュータの利用登録を行い、DDBJ に含まれる 1000 人ゲノムプロジェクトの 26 ヒト集団、約 2500 人以上の全ゲノム塩基配列データから、縦列反復配列データを抽出できる GangSTR や LobSTR などのソフトウェアを利用して、日本人で識別力の高い 4 塩基リピートの 218 座位(文献)の STRs の遺伝型の抽出を試みた。

(4) 東京大学人類遺伝学研究室から提供された、日本人約 400 人及び韓国人約 90 人の全ゲノム塩基配列データから in silico で得られた日本人で約 5000 箇所と韓国人で約 3000 箇所の Alu 挿入部位のデータの評価を行った後、日本人と韓国人(韓族)で挿入箇所が異なり、かつ、それぞれの集団で 20%以上の頻度を持つ挿入部位を選び、各集団に由来することだけが既知の各 8 名の DNA 試料について、予備的にその挿入箇所データの評価を行った。

(5) 新規 Alu 探索法であるレトロトランスポゾン・ディスプレイ法について、次世代シーケンサ MiSeq による解析までの検討を行った。

(6) 最近注目されているマイクロハプロタイプ (MH) について、1000 人ゲノムプロジェクトのデータから今まで知られている 418 マーカー(文献)を、日本人(JPT)105 人及び中国人(CHB) 102 人から抽出し、両集団の各 MH のハプロタイプ頻度を算出し、ヘテロ接合度をもとめた。また、両集団間の各 MH のアレル頻度分布の有意差検定を行った。

4. 研究成果

(1) 105 座位の STR のデータを利用して、統計学的に判別可能かどうか検討した結果、ベイズ的アプローチにより、強い証拠とされるベイズ因子(BF)が 10 以上になる日本人は約 92%となり、他方、韓族であると考えられる BF が 0.1 未満の日本人は約 28%であった。このことから、日本人由来であるというのは容易であるが、韓族由来であるというのは難しいという結果が得られた。その結果は、論文として外国誌に掲載された。

(2) 日本人及び中国人 32 名の DNA 試料から得られた、信頼できる 261 座位のデータについて、得られた結果は、以下のとおりである。

両集団の各座位について、アレル頻度を算出し、ヘテロ接合度などの法科学的統計量を算出した。その結果、日本人及び中国人で観察された平均アレル数はそれぞれ 6.65 及び 6.56 で、ヘテロ接合度(観察値)の平均値は、それぞれ 0.72 及び 0.73 で、ほぼ同じであった。両ヒト集団で HWE からの乖離が見られなかった座位数は、 2 検定で、201、Exact テストで 218 で、両検定共乖離が見られなかったのは、185 座位であった。(Table 1)

HWE からの乖離が見られなかった 185 座位について、それぞれ両集団間のアレル頻度分布の差を見たところ、有意な差 ($P < 0.05$) が見られたのは、14 座位のみであった。(Table 1)

両集団について、集団遺伝学関係や異質性を見るために、主成分分析、因子対応分析、総尤度比の算出を行ったところ、おおよそ各常染色体の短腕・長腕に位置する 45 座位程度の比較的識別力の高い STRs をマーカーとすれば、両集団の識別が可能であろうことが示唆された。(図 1)

これらの結果は、国際学会で発表され、論文として掲載された。

今後は、これらのデータをもとに確率論的な計算方法を検討する必要がある。また、将来的に Long-Read の NGS と個人インデックス配列を組み合わせれば、より効率的に型判定が可能になると考えられる。

Table 1. The relationship between the numbers of deviation from HWE and P-values in each population and both populations.

P-value	GenAIEx 6.503 software (χ^2 test)		Arlequin 3.5.2.2 software (Exact test)	
	Population (Pop)		Population (Pop)	
	Japanese	Chinese	Japanese	Chinese
P>0.05	236	241	P>0.05	210 216
P<0.05	16	14	P<0.05	15 11
P<0.01	7	4	P<0.01	11 11
P<0.001	2	2	P<0.001	11 8
Total	261	261	Total	261 261
Both pops: P>0.05		201	Both pops: P>0.05 218	
Both tests P>0.05		185		

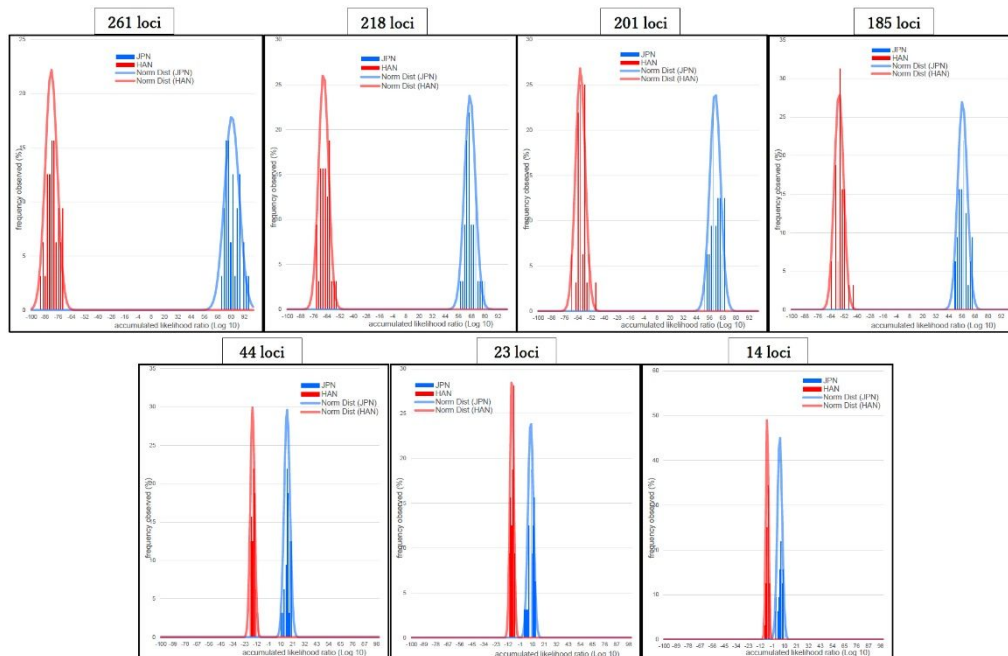


図1. 各座位における総尤度比の分布

(3) DDBJに含まれる1000人ゲノムプロジェクトから、日本人で識別力の高い4塩基リピートの218座位(文献)のSTRsの遺伝型の抽出をGangSTRにより試みた。データ量の点から、一度に10人ずつ、218座位のリピート配列を抽出したところ、約60座位でほぼ正確と考えられる塩基配列から得られるリピート数を推定することが可能であった。追加的に抽出した日本人20名及びヨーロッパ人10名からも同様の結果が得られた。しかしながら、再現性を求め重複解析を行ったところ、それらの結果は正確ではないことが判明した。また、LobSTRは、WEB上から消失したので、in silicoでのSTR解析は中断せざるを得なかった。

(4) In silicoで得られた日本人で約5000箇所と韓国人で約3000箇所のAlu挿入箇所を比較・評価したところ、両集団間で挿入箇所が異なり、比較的新しい挿入Alu配列のサブタイプであるAlu-Yタイプがあまり多くないことが分かった。

(5) 当初より計画されていたレトロトランスポゾン・ディスプレイ法と次世代シーケンサMiSeqとを併用して、in vitroでの分散型反復配列のAlu配列挿入箇所の検出は、植物などのゲノムに比して、ヒトゲノムが大きいこと、ゲノムDNAをgTUBEなどで機械的に効率よく断片化することが非常に難しいことが判明した。

(6) 日本人においてSNPsが2~4個ある259座位のMHについて、日本人(JPT)105人及び中国人(CHB)102人のハプロタイピングを行った。両集団の各MHのハプロタイプ頻度を算出し、ヘテロ接合度をもとめたところ、各SNPs数におけるJPT及びCHBのヘテロ接合度の平均値は、全て約0.5以下で、SNPsが多いマーカーほど高い値を示す傾向が見られた。また、両集団間の各MHのアレル頻度分布の有意差検定を行ったが、差はほとんど見られなかった。従って、MHは、遺伝的に近いと考えられる集団や個人の識別力はSTRよりも低いことが示唆された。

<引用文献>

- Satoshi Hirata, Kaname Kojima, Kazuharu Misawa, Olivier Gervais, Yosuke Kawai, Masao Nagasaki. Population-scale whole genome sequencing identifies 271 highly polymorphic short tandem repeats from Japanese population. *Heliyon* 4 (2018) e00625. doi: 10.1016/j.heliyon.2018. e00625
- Daniel S. Standage and Rebecca N. Mitchell. MicroHapDB: A Portable and Extensible Database of All Published Microhaplotype Marker and Frequency Data. *Frontiers in Genetics* 11 (2020) 781. doi: 10.3389/fgene.2020.00781

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 4件/うち国際共著 1件/うちオープンアクセス 0件）

1. 著者名 Toshimichi Yamamoto, Kentaro Doi, Ryutaro Fukami, Takashi Yoshimoto, Yujin Natori, Akira Ishi	4. 巻 8
2. 論文標題 Comparisons between Japanese and Han Chinese populations for 261 autosomal STR loci	5. 発行年 2022年
3. 雑誌名 Forensic Science International: Genetics Supplement Series	6. 最初と最後の頁 117-119
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.fsigs.2022.10.003	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 佐藤 史織, 李 知娟, 佐野 開人, 吉本 高士, 石井 晃, 河合 洋介, 徳永 勝士, 山本敏充	4. 巻 28
2. 論文標題 日本人の全ゲノム塩基配列データから in silico で得られた Alu 配列挿入箇所の評価	5. 発行年 2020年
3. 雑誌名 DNA多型	6. 最初と最後の頁 125-128
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 無
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 Koji Tsukuda, Shuhei Mano, Toshimichi Yamamoto	4. 巻 -
2. 論文標題 Bayesian approach to discriminant problems for count data with application to multilocus short tandem repeat dataset	5. 発行年 2020年
3. 雑誌名 Statistical Applications in Genetics and Molecular Biology	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） 10.1515/sagmb-2018-0044	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 荒川智哉, 吉本高士, 佐藤史織, 石井晃, 山本敏充	4. 巻 27
2. 論文標題 エルサルバドルにおける17座位のY-STRs及びY-hgsの法科学的及び集団遺伝学的解析	5. 発行年 2019年
3. 雑誌名 DNA多型	6. 最初と最後の頁 177-180
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 無
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 Isao Yuasa, Atsushi Akane, Toshimichi Yamamoto, Aya Matsusue, Minoru Endoh, Mayumi Nakagawa, Kazuo Umetsu, Takaki Ishikawa, Morio Iino	4. 巻 33
2. 論文標題 Japaneseplex: A forensic SNP assay for identification of Japanese people using Japanese-specific alleles	5. 発行年 2018年
3. 雑誌名 Legal Medicine	6. 最初と最後の頁 17-22
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.legalmed.2018.04.008	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 Hiroaki Nakanishi, Vania Pereira, Claus Boslarshting, Toshimichi Yamamoto, Torben Tvedebrink, Masaaki Hara, Aya Takada, Kazuyuki Saito, Niels Morling	4. 巻 33
2. 論文標題 Analysis of mainland Japanese and Okinawan Japanese populations using the precision ID Ancestry Panel	5. 発行年 2018年
3. 雑誌名 Forensic Science International: Genetics	6. 最初と最後の頁 106-109
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.fsigen.2017.12.004	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 山本敏充, 柘植雅大, 小川久恵, 吉本高士, 石井晃	4. 巻 27
2. 論文標題 エルサルバドルの少数民族ピピル族における15座位のSTR解析	5. 発行年 2018年
3. 雑誌名 DNA多型	6. 最初と最後の頁 164-167
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 無
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 山本敏充, 浅井一, 廣重優二, 小川久恵, 吉本高士, 石井晃	4. 巻 26
2. 論文標題 インドネシアにおける15 STRsのアレル頻度分布及び世界中のヒト集団との集団遺伝学的解析	5. 発行年 2017年
3. 雑誌名 DNA多型	6. 最初と最後の頁 138-142
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 無
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8件（うち招待講演 0件 / うち国際学会 4件）

1. 発表者名 山本敏充, 土井健太郎, 吉本高士, 名取雄人, 石井晃
2. 発表標題 常染色体上の261座位のSTRsデータにおける日本人と中国人の法科学・集団遺伝学的解析
3. 学会等名 第107次日本法医学会学術全国集会
4. 発表年 2023年

1. 発表者名 Toshimichi Yamamoto, Kentaro Doi, Ryutarō Fukami, Takashi Yoshimoto, Yujin Natori, Akira Ishii
2. 発表標題 Comparisons between Japanese and Han-Chinese populations in terms of 261 autosomal STR loci
3. 学会等名 The 29th Congress of the International Society for Forensic Genetics (ISFG 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Toshimichi Yamamoto, Kaito Sano, Ji yeon Lee, Shiori Sato, Takashi Yoshimoto, Akira Ishii, Yosuke Kawai, Katsushi Tokunaga
2. 発表標題 Evaluation of Alu insertion positions observed in silico in Japanese whole genome sequencings
3. 学会等名 International Society of Forensic Genetics (ISFG) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 佐藤史織, 李知娟, 佐野開人, 吉本高士, 石井晃, 河合洋介, 徳永勝士, 山本敏充
2. 発表標題 日本人の全ゲノム塩基配列データから in silico で得られた Alu 配列挿入箇所の評価
3. 学会等名 日本 DNA 多型学会第28回学術集会
4. 発表年 2019年

1. 発表者名 Toshimichi Yamamoto, Masahiro Tsuge, Takashi Yoshimoto, Akira Ishii
2. 発表標題 Forensic and Population Genetic Analyses for 15 STR Loci among Three Regional Populations of Small Ethnic Group “Pipil” in El Salvador
3. 学会等名 International Society of Forensic Genetics-English Speaking Working Group (ISFG-ESWG) 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 荒川智哉, 吉本高士, 佐藤史織, 石井晃, 山本敏充
2. 発表標題 エルサルバドルにおける17座位のY-STRs及びY-hgsの法科学的及び集団遺伝学的解析
3. 学会等名 日本DNA多型学会第27回学術集会
4. 発表年 2018年

1. 発表者名 Hiroaki Nakanishi, Vania Pereira, Claus Boslshrsting, Toshimichi Yamamoto, Torben Tvedebrink, Masaaki Hara, Aya Takada, Kazuyuki Saito, Niels Morling
2. 発表標題 Ancestry analysis of mainland Japanese and Okinawa Japanese populations using the Precision ID Ancestry Panel
3. 学会等名 International Society of Forensic Genetics (ISFG) 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 山本敏充, 柘植雅大, 小川久恵, 吉本高士, 石井晃
2. 発表標題 エルサルバドルの少数民族ピピル族における15座位のSTR解析
3. 学会等名 日本DNA多型学会第26回学術集会
4. 発表年 2018年

〔図書〕 計1件

1. 著者名 井ノ上 逸朗, 今西 規, 河村 正二, 斎藤 成也, 颯田 葉子, 田嶋 敦	4. 発行年 2021年
2. 出版社 一色出版	5. 総ページ数 448
3. 書名 ヒトゲノム事典	

〔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
研究協力者	間野 修平 (MANO Shuhei)		
研究協力者	佃 康司 (TSUKUDA Koji)		

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