

令和 5 年 6 月 26 日現在

機関番号：63801

研究種目：挑戦的研究(萌芽)

研究期間：2020～2022

課題番号：20K21405

研究課題名(和文) CRISPRを利用した新規ウイルス(ファージ)ゲノム同定そして進化の考察

研究課題名(英文) Identification of viruses using CRISPR spacer sequences.

研究代表者

井ノ上 逸朗(Inoue, Ituro)

国立遺伝学研究所・ゲノム・進化研究系・特任教授

研究者番号：00192500

交付決定額(研究期間全体)：(直接経費) 5,000,000円

研究成果の概要(和文)：ウイルスは地球上でもっとも多くの遺伝物質を有する有機体である。全てのウイルスは特定の細胞に感染し増殖する。多くのウイルスのゲノムを解析することで、ウイルスのみならず細胞性生命の進化と起源についての新たな理解も得られる。これまでのウイルスゲノムを探索する研究は、相同性検索によって検出される、既知のウイルスに近縁な種に限られていた。本研究の目的は、1)CRISPR免疫記憶を利用して相同性検索に依存しない方法でファージゲノムを探索し、さらに得られたゲノム配列を用いてウイルス全体の進化と起源を明らかにすること、2) 古代ウイルス配列決定をおこない、ウイルス進化を年代確立した上で検討すること、である。

研究成果の学術的意義や社会的意義

本研究では得られるメタゲノムデータからCRISPR-Casシステムを用いホモロジーに頼らないウイルス同定を確立できた。縄文人をはじめ多数の年代の日本列島人の歯髄と歯石から得られた76の全ゲノムデータと縄文人糞石由来DNA4検体の提供を受け、全ゲノムデータを取得した。縄文人から得られた歯髄および糞石由来の全ゲノムデータを用い、ウイルスの網羅的探索を行った。ウイルス同定のために、データベース上に登録のある既知のウイルスを参照配列として相同性検索を行い、11種類の現代のウイルスと類似のゲノムを持つ縄文ウイルスを同定した。

研究成果の概要(英文)：Viruses are the most abundant population of organisms on earth that possess the most abundant genetic materials. All viruses infect and multiply in specific cells. Horizontal gene spread and arms race evolutionary competition between viruses and cells have been the driving forces of evolution. Analysis of the genomes of many viruses will provide insight into the evolution and origin of not only viruses but also cellular life. Previous studies exploring viral genomes have been limited to species closely related to known viruses, as detected by homology searches. The objectives of this study are (1) to search phage genomes using CRISPR immuno-memory in a manner that does not rely on homology search, and to use the obtained genome sequences to elucidate the evolution and origin of entire viruses, (2) To examine ancient viral sequences to establish the chronology of viral evolution.

研究分野：人類遺伝学, ゲノム医科学

キーワード：新規ウイルス CRISPR-Cas ファージ 縄文人

## 1. 研究開始当初の背景

我々は2017年にヒトのレトロウイルスデータベース(<http://herv-tfbs.com>)を構築した(Ito et al, PLOS Genet 2017, <https://doi.org/10.1371/journal.pgen.1006883>)。この研究はレトロウイルスの持つ転写因子結合部位がヒトの遺伝子発現ネットワークに与える影響を調査する上での基礎を構築するものである。以降、我々はヒトの遺伝子発現のみならず、細胞性生命のタンパクの進化に与えたウイルスの影響を調査しようと試みている。ウイルスと細胞で共有するタンパクドメインの検出と、分子系統樹解析を行うには広汎かつ大量のシーケンスデータが必要である。

## 2. 研究の目的

地球上にはおおよそ  $10^{31}$  個のウイルスが存在すると推定されている。これは細菌総数のさらに1桁上である。ウイルスは膨大かつ多様な遺伝情報を有する有機体として捉えることができる。これまでのウイルスゲノムを探索する研究は、相同性検索によって検出される、既知のウイルスに近縁な種に限られていた。本研究の目的は、CRISPR 免疫記憶を利用して相同性検索に依存しない方法でファージゲノムを探索し、さらに得られたゲノムを用いてウイルス全体の進化と起源を明らかにすることである。ウイルスゲノムには多様かつ未知の遺伝子が大量にコードされており、その大部分は細胞性生命で見つかる遺伝子とはホモロジーがない。生物の進化は遺伝子で測られる。大量の未知遺伝子を持つウイルスゲノムを網羅的に解析することは、生物学の根源的なテーマである生命の起源を解き明かす上で重要な役割があると我々は考える。ウイルスゲノムを解析することで得られる情報はウイルスのみの進化に関するものに限られない。全てのウイルスは特定の細胞に感染し、その細胞の複製機構を流用することで増殖する。ウイルスと細胞の間では遺伝子の水平伝播と進化的軍拡競争が起こり、これらは生物進化のひとつの要因となった。例えば真核生物のテロメラーゼはレトロウイルスやグループ II イントロンの持つ逆転写酵素から進化したと推定されているし、細菌由来とされるミトコンドリア DNA ポリメラーゼはファージの DNA ポリメラーゼ と高い相同性を持つことが知られている。あらゆる環境から得たウイルスゲノムを広汎に解析することで、ウイルスのみならず、細胞性生命の進化と起源についての深い理解も得られると我々は考える。

## 3. 研究の方法

次世代シーケンサーの発展によって、大規模なメタゲノムシーケンシングが可能となった。メタゲノムとは、ある環境中に存在する細菌、古細菌、真核生物そしてウイルス全ての遺伝情報である。海水や土壌、糞便などから DNA を取り出し、次世代シーケンサーで読み取ることで、その環境中に存在する生物あるいはウイルスの遺伝情報を得ることができる。メタゲノムから得たあるシーケンスがウイルス由来であることを示すには次の挑戦課題を克服せねばならなかった。

1. ウイルスゲノムは非常に多様である
2. ウイルス遺伝子の多くは他の生物との相同性がほとんどない
3. ウイルスはリボソーム RNA のようなマーカー遺伝子を持たない
4. ウイルスのリファレンスゲノムは非常に限られている

既存の方法では、既知のウイルスゲノムで見つかるタンパクに相同なタンパクをコードするコンティグをメタゲノムから検出していた。この方法は既知のウイルス種に近縁なウイルスゲノムを効率よく検出できる。しかし、ウイルスゲノムにコードされる遺伝子は非常に多様であり、種分化の起源についても未解明の部分が多い。したがって、この方法は種分化の起源が古いウイルス、あるいは全く新規の遺伝子を持つウイルスを検出するには不十分であると我々は考えた。

これを解決するために、我々は CRISPR 免疫記憶を利用した方法を提案した。原核生物に感染するウイルスはファージと呼ばれる。原核生物はファージに対抗する手段の一つとして CRISPR-Cas システムを持つ。CRISPR-Cas システムは原核生物における獲得免疫機構であり、特に CRISPR は獲得免疫記憶として働く。CRISPR 遺伝子座には過去に細胞に侵入したファージゲノムの一部が記憶として保存されている。したがって CRISPR にコードされるファージゲノムの断片を用いることで、タンパクの相同性を用いることなく、ファージゲノムをメタゲノムから検出することができる。

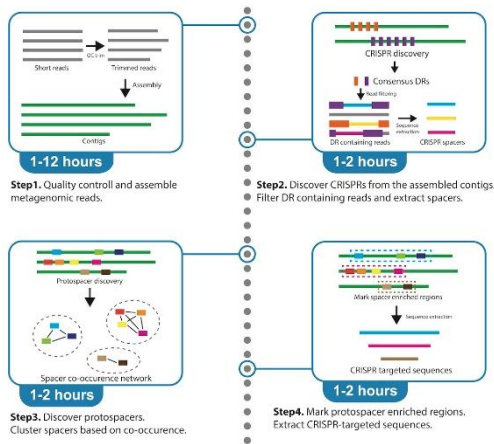


図 1. CRISPR を使ったファージゲノム検出法

115,645 のコンティグを同定した。この内、1,016 のコンティグは末端重複、かつカプシド遺伝子をコードしており、ファージ全長ゲノムと推定される。さらに、これらのファージ全長ゲノムのうち 298 は既存のデータベースに登録されていない新規ゲノムであった。

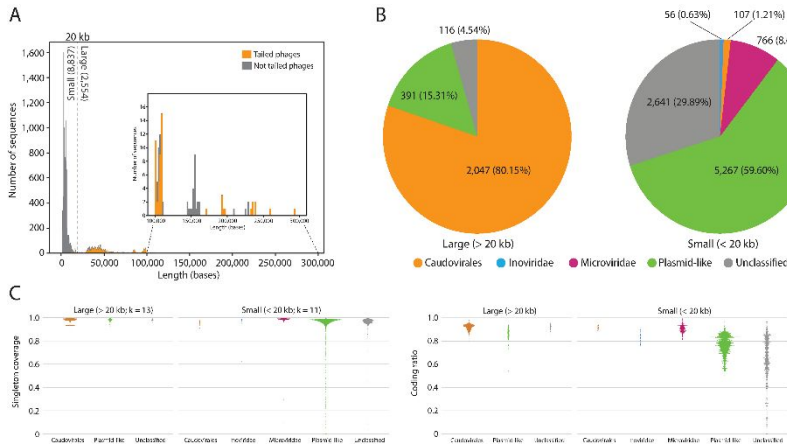


図 2. 検出されたファージゲノムの全長の分布

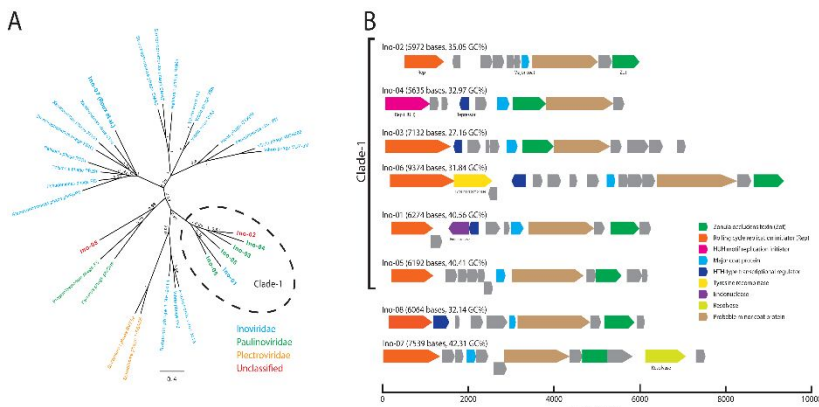


図 3. 検出された 10kbp 以下のファージゲノムの距離に基づいた系統樹

図 1 に解析の概要を示す。メタゲノムのショートリードシーケンスをアセンブリし、長いシーケンス（コンティグ）を得る。得られたコンティグから CRISPR 遺伝子座を検出し、そこからファージ由来の配列、いわゆるスペーサー配列を抽出する。集積したスペーサー配列をメタゲノムコンティグにアラインメントする。複数のスペーサーがヒットしたコンティグはファージ由来である可能性がある。さらに遺伝子推定と相同性検索を用いてこれらのコンティグに含まれる遺伝情報を詳細に調べた。事前調査として、我々は NCBI に公開されている 511 ランのメタゲノムシーケンス (9.8Tbp) を解析し、上記の方法を適用した。結果、アセンブリしたコンティグから 463,248 のユニークな CRISPR スペーサー配列を抽出し、最低 4 つのスペーサーによってターゲットされる

図 2 は検出されたゲノムの全長の分布とヒットしたデータベースの内訳である。IMG\_VR は先述したタンパクの相同性を用いて検出されたウィルスゲノムのデータベースである。検出された長さが 10kbp から 200kbp のゲノムの大部分は IMG\_VR に登録済みである。10kbp 以下の小さいゲノムのほぼ半数は NCBI に登録済み、そして残りは新規である。250kbp 以上の大きいゲノムの全ては新規である。

図 3 に 10kbp 以下のゲノムの距離に基づいた系統樹を示す。既知のウィルスは系統樹上の互いに近い位置に現れている。我々の方法は複数の新規ファージランチを検出できている。これらの結果は我々の方法が既知のファージに近縁な種に限らない、特に極端に大きいか小さいファージゲノムを検出可能であることを示している。本研究において、我々は上記の方法

を広汎なメタゲノムに適用し、さらなるファージゲノムの集積を行った。解析の精度を向上し、非末端重複のゲノム、プロファージ、そして RNA ファージの検出も行った。さらに、得られたファージゲノムとそこから推定されるウィルスタンパクのデータベースを構築した。ウィルスタンパクの隠れマルコフモデルを計算し、これを用いてウィルスと細胞性生命で共有するタンパクドメインを検出した。細胞性生命の全タンパクファミリーのうち、ウィルスが進化に関わったと推定されるタンパクの数を定量的に評価した。これらのタンパクの分子系統樹から、ウィルスと細胞の間で起こった遺伝子の水平伝播の時期を推定した。また、タンパクの機能と分布からそのタンパクファミリーの起源を推定した。細胞性生命の進化におけるウィルスの関わりを明らかにし、ウィルスと生命の起源について考察した。

#### 4. 研究成果

ウイルスは地球上で最も多様な存在であり、未知のゲノム情報がそこにはある。ウイルスゲノムには多様性かつ大量の遺伝子がコードされており、その大部分は細胞性生命で見つかる遺伝子とはホモロジーがない。大量の未知遺伝子を持つウイルスゲノムを網羅的に解析することは、生物学の根源的なテーマである生命の起源を解き明かす上で重要な役割がある。ウイルスゲノムを解析することで得られる情報はウイルスのみの進化に関するものに限られない。

本研究では得られるメタゲノムデータから CRISPR-Cas システムを用いホモロジーに頼らないウイルス同定を確立できた。さらには縄文人検体から古代ウイルス同定をおこなった。古代ウイルスは、古代人の口や腸に存在していたと推定される。そのため、歯髄や糞石といったサンプルには縄文人のゲノムに加え、ウイルスのゲノムも存在していたと考えられる。これらのサンプルから全ゲノムシーケンシングデータ(全ゲノムデータ) を得ることを着想した。国立科学博物館の篠田謙一館長より縄文人をはじめ多数の年代の日本列島人の歯髄と歯石から得られた 76 の全ゲノムデータの提供を受けた。加えて東京大学理学部太田博樹教授より、縄文人糞石由来 DNA4 検体の提供を受け、全ゲノムデータを取得した。さらに、日本と海外の古代人のバイロームを比較できるように、データベース上で公開されている海外の古代人由来全ゲノムデータもダウンロードした。こうして得られた古代人由来の全ゲノムデータを使うことで古代ウイルス情報を得ることができる。縄文人から得られた歯髄および糞石由来の全ゲノムデータを用い、ウイルスの網羅的探索を行った。ウイルス同定のために、データベース上に登録のある既知のウイルスを参照配列として相同性検索を行った。これによって 11 種類の現代のウイルスと類似のゲノムを持つ縄文ウイルスを同定した。

## 5. 主な発表論文等

〔雑誌論文〕 計4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 3件）

1. 著者名 Sugimoto Ryota, Nishimura Luca, Nguyen Phuong Thanh, Inoue Ituro	4. 巻 3
2. 論文標題 Extraction of CRISPR-targeted sequences from the metagenome	5. 発行年 2022年
3. 雑誌名 STAR Protocols	6. 最初と最後の頁 101525 ~ 101525
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.xpro.2022.101525	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 Nishimura Luca, Fujito Naoko, Sugimoto Ryota, Inoue Ituro	4. 巻 14
2. 論文標題 Detection of Ancient Viruses and Long-Term Viral Evolution	5. 発行年 2022年
3. 雑誌名 Viruses	6. 最初と最後の頁 1336 ~ 1336
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/v14061336	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 Ryota Sugimoto, Luca Nishimura, Phuong Thanh Nguyen, Jumpei Ito, Nicholas F Parrish, Hiroshi Mori, Ken Kurokawa, Hirofumi Nakaoka, Ituro Inoue	4. 巻 17
2. 論文標題 Comprehensive discovery of CRISPR-targeted terminally redundant sequences in the human gut metagenome: Viruses, plasmids, and more	5. 発行年 2021年
3. 雑誌名 PLoS computational 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) 10.1371/journal.pcbi.1009428	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 Luca Nishimura, Ryota Sugimoto, Jun Inoue, Hirofumi Nakaoka, Hideaki Kanzawa-Kiriyama, Ken-Ichi Shinoda, Ituro Inoue	4. 巻 66
2. 論文標題 Identification of ancient viruses from metagenomic data of the Jomon people	5. 発行年 2021年
3. 雑誌名 Journal of human genetics	6. 最初と最後の頁 287-296
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s10038-020-00841-6	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 -

〔学会発表〕 計22件（うち招待講演 0件 / うち国際学会 7件）

1. 発表者名 Luca Nishimura, Ryota Sugimoto, Inoue Ituro
2. 発表標題 Virome analyses of the ancient individuals who lived in the Japanese archipelago 3,000 years ago
3. 学会等名 International Virus Bioinformatics Meeting 2022 (国際学会) (国際学会)
4. 発表年 2022年

1. 発表者名 Ryota Sugimoto, Luca Nishimura, Phuong Thanh Nguyen, Jumpei Ito, Nicholas F. Parrish, Hiroshi Mori, Ken Kurokawa, Hirofumi Nakaoka and Ituro Inoue
2. 発表標題 Comprehensive discovery of CRISPR targeted sequences from the human gut metagenome
3. 学会等名 International Virus Bioinformatics Meeting 2022 (国際学会) (国際学会)
4. 発表年 2022年

1. 発表者名 杉本竜太、西村瑠花、井ノ上逸朗
2. 発表標題 ヒト腸内メタトランスクリプトームから検出されたRNAウイルス
3. 学会等名 日本進化学会第24回年会
4. 発表年 2022年

1. 発表者名 Ryota Sugimoto, Luca Nishimura, Inoue Ituro
2. 発表標題 Comprehensive discovery of CRISPR-targeted sequences in the human gut metagenome
3. 学会等名 2022 annual meeting of the American Society of Human Genetics (国際学会) (国際学会)
4. 発表年 2022年

1. 発表者名 杉本竜太、西村瑠佳、井ノ上逸朗
2. 発表標題 CRISPR-targeted non-transcribed RNA sequences in the human gut metatranscriptome
3. 学会等名 日本分子生物学会 第45回年会
4. 発表年 2022年

1. 発表者名 西村瑠佳, 杉本竜太, 神澤秀明, 篠田謙一, 井ノ上逸朗
2. 発表標題 日本列島古代人のゲノムデータを用いた古代微生物解析
3. 学会等名 日本進化学会第24回大会
4. 発表年 2022年

1. 発表者名 Luca Nishimura, Ryota Sugimoto, Hideaki Kanzawa-Kiriyama, Ken-ichi Shinoda, Ituro Inoue
2. 発表標題 Ancient virome analyses of ancient individuals who lived in Japanese archipelago
3. 学会等名 ASHG Annual meeting 2022 of American Society of Human Genetics (国際学会)
4. 発表年 2022年

1. 発表者名 西村瑠佳, 杉本竜太, 神澤秀明, 篠田謙一, 井ノ上逸朗
2. 発表標題 古代人由来全ゲノムシーケンシングデータを用いた古代ウイルス探索
3. 学会等名 日本分子生物学会第45回年会
4. 発表年 2022年

1. 発表者名 西村瑠佳、杉本竜太、井ノ上逸朗
2. 発表標題 古代人由来全ゲノムシーケンシングデータを用いた古代バイローム解析
3. 学会等名 日本人類遺伝学会 第67回大会
4. 発表年 2022年

1. 発表者名 西村瑠佳、谷野彰勇、鯉本真友美、稲田健一、勝村啓史、小川元之、小金淵佳江、和久大介、熊谷真彦、杉本竜太、太田博樹、井ノ上逸朗
2. 発表標題 縄文人由来DNAを用いた古代微生物の配列解析
3. 学会等名 日本進化学会 第23回大会
4. 発表年 2021年

1. 発表者名 杉本竜太、西村瑠佳、THANH Phuong Nguyen、伊東潤平、PARRISH Nicholas F.、森宙史、黒川顕、中岡博史、井ノ上逸朗
2. 発表標題 CRISPRにターゲットされるヒト腸内メタゲノム配列の網羅的同定
3. 学会等名 日本進化学会 第23回大会
4. 発表年 2021年

1. 発表者名 Luca Nishimura, Akio Tanino, Takafumi Katsumura, Kae Koganebuchi, Daisuke Waku, Masahiko Kumagai, Ryota Sugimoto, Naoko Fujito, Hiroki Oota, Ituro Inoue
2. 発表標題 Gut microbiome analyses of ancient individuals, so called "Jomon", lived in Japanese archipelago
3. 学会等名 ASHG Virtual Meeting 2021 (国際学会)
4. 発表年 2021年



1. 発表者名 西村瑠佳、谷野彰勇、鯨本真友美、稲田健一、勝村啓史、小川元之、小金淵佳江、和久大介、熊谷真彦、杉本竜太、藤戸尚子、太田博樹、井ノ上逸朗
2. 発表標題 縄文人糞石由来の古代DNAを用いた微生物解析
3. 学会等名 日本分子生物学会 第44回年会
4. 発表年 2021年

1. 発表者名 杉本竜太、西村瑠佳、THANH Phuong Nguyen、伊東潤平、PARRISH Nicholas F.、森宙史、黒川顕、中岡博史、井ノ上逸朗
2. 発表標題 CRISPRにターゲットされるヒト腸内メタゲノム配列の網羅的検出
3. 学会等名 日本分子生物学会 第44回年会
4. 発表年 2021年

1. 発表者名 西村瑠佳
2. 発表標題 生命情報科学への新規参入とウイルスゲノム解析
3. 学会等名 日本バイオインフォマティクス学会年会
4. 発表年 2020年

1. 発表者名 西村瑠佳、杉本竜太、井上潤、中岡博史、神澤秀明、篠田謙一、井ノ上逸朗
2. 発表標題 縄文人由来DNAを用いた古代微生物の配列解析
3. 学会等名 日本進化学会 第22回大会
4. 発表年 2020年

1. 発表者名 杉本竜太、西村瑠佳、井ノ上逸朗
2. 発表標題 CRISPR免疫記憶を使った新規ウイルス同定
3. 学会等名 日本進化学会 第22回大会
4. 発表年 2020年

1. 発表者名 Luca Nishimura, Ryota Sugimoto, Jun Inoue, Hirofumi Nakaoka, Hideaki Kanzawa, Kenichi Shinoda, Ituro Inoue
2. 発表標題 Metagenomic analyses of viruses from ancient individuals of 3,000 years ago lived in Japanese archipelago
3. 学会等名 ASHG Virtual meeting 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Ryota Sugimoto, Luca Nishimura, Ituro Inoue
2. 発表標題 Detection of viral genomes from human gut metagenome using CRISPR adaptive immunological memory
3. 学会等名 ASHG Virtual meeting 2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Luca Nishimura, Ryota Sugimoto, Jun Inoue, Hirofumi Nakaoka, Hideaki Kanzawa, Kenichi Shinoda, Ituro Inoue
2. 発表標題 Metagenomic analyses of viruses from ancient DNA of Jomon people
3. 学会等名 日本人類遺伝学会 第65回大会
4. 発表年 2020年

1. 発表者名 西村瑠佳、杉本竜太、井上潤、中岡博史、神澤秀明、篠田謙一、井ノ上逸朗
2. 発表標題 縄文人由来DNAを用いた古代微生物の配列解析
3. 学会等名 日本分子生物学会 第43回年会
4. 発表年 2020年

1. 発表者名 杉本竜太、西村瑠佳、THANH Phuong Nguyen、伊東潤平、PARRISH Nicholas F.、森宙史、黒川顕、中岡博史、神澤秀明、篠田謙一、井ノ上逸朗
2. 発表標題 CRISPR免疫記憶を使った新規ウイルス検出
3. 学会等名 日本分子生物学会 第43回年会
4. 発表年 2020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
---------	---------