

令和 3 年 6 月 17 日現在

機関番号：13101

研究種目：基盤研究(C) (一般)

研究期間：2017～2020

課題番号：17K00401

研究課題名(和文) 水平伝播遺伝子予測システムの開発と環境適応と共進化過程の解明

研究課題名(英文) Development of horizontal gene transfer detection system and elucidation of environmental adaptation and co-evolution process

研究代表者

阿部 貴志 (Abe, Takashi)

新潟大学・自然科学系・教授

研究者番号：30390628

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：水平伝播遺伝子探索はアミノ酸配列の類似性検索による探索がなされてきたが、一部の遺伝子のみが検出可能であった。ゲノム上の全遺伝子を対象とした網羅的検出手法が必要である。連続塩基組成のみに着目してゲノム配列断片を生物種別に高精度に分類可能な一括学習型自己組織化マップによる水平伝播の候補遺伝子検出と由来生物系統の推定手法を開発した。南極コケ坊主由来の *Sphingomonas* 属細菌2種での解析にて、水平伝播の由来は大きく異なったが、獲得した遺伝子機能に共通性が示唆された。開発手法は、HGT候補とその起源をゲノム全体で検出するだけでなく、微生物の環境適応への新しい視点を提供できる。

研究成果の学術的意義や社会的意義

生物の進化や環境適応には生物種間での遺伝子の水平伝播が大きな役割を果たしている。配列相同性検索などの従来手法では水平伝播遺伝子の検出精度に限界があったが、本研究にて開発した水平伝播予測ワークフローは、ゲノム全体における水平伝播遺伝子の全容解明を可能とした。南極由来細菌の異なる系統の同属異種間における水平伝播遺伝子探索を行ったところ、水平伝播で獲得した遺伝子の機能やアミノ酸組成の変化など多くの共通した特徴が見られ、南極環境に適応するプロセスにおいて収斂進化の可能性が示唆された。様々な環境下での生育環境適応戦略や共生関係が及ぼしてきた共進化過程の解明への活用が期待できる。

研究成果の概要(英文)：Horizontal gene transfer (HGT) has been searched by amino acid sequence similarity search, but only some genes could be detected. A comprehensive HGTs detection method for all genes on the genome is required. Focusing only on the oligonucleotide composition, we have developed a method for detecting candidate genes for HGTs and estimating its origin using a batch learning self-organizing map that can classify genome sequence fragments by biological type with high accuracy. As an application, analysis of two species of *Sphingomonas* spp. derived from Antarctic moss showed that the origin of HGT gene was significantly different, but the HGT gene functions were common. The development method is useful not only for detecting HGT candidates and their origins throughout the genome, but also as a method that can provide a new perspective on the environmental adaptation of microorganisms.

研究分野：バイオインフォマティクス

キーワード：一括学習型自己組織化マップ 連続塩基組成 水平伝播遺伝子 環境適応 共生細菌

1. 研究開始当初の背景

近年、極限環境や共生関係下の生態系の解明を含む、様々な観点での環境微生物のゲノム解読プロジェクトが爆発的に増え続けている。生物種間での遺伝子の水平伝播が生物進化や環境適応に大きな役割を果たしてきたことは定説となっているが、受けて側のゲノムに埋め込まれた水平伝播遺伝子は長い進化の過程で痕跡化しつつある。従来の相同性検索では、高精度かつ網羅的に水平伝播遺伝子とその送り出し側の由来種を特定することが困難であり、相同性検索とは異なった原理に基づく超大規模データ解析法の確立が求められている。一括学習型自己組織化マップ (BLSOM) は連続塩基組成のみで生物種ごとに高精度に分離できるが、水平伝播遺伝子は各生物種のゲノムに固有の GC 含量やコドン組成などと異なることが多く、むしろその由来種 (送り出し側) に近い特徴を示している。網羅的かつ高精度に由来種を予測することができれば、生育環境適応戦略や共生関係が及ぼしてきた共進化過程の解明に繋げることが可能となりえる。

2. 研究の目的

生物の進化や環境適応には生物種間での遺伝子の水平伝播が大きな役割を果たしている。しかしながら、従来の配列相同性検索を用いた水平伝播遺伝子の検出は限界があり、ゲノム全域を対象にした水平伝播遺伝子の全容解明には新規解析手法の開発が求められている。我々が開発した BLSOM は、配列相同性に依らずに、連続塩基組成のみで生物種ごとに高精度に分離できることから、着目生物種に固有の GC 含量やコドン組成とは異なる視点で、水平伝播遺伝子の検出が期待できる。本申請は BLSOM を活用した水平伝播遺伝子候補の予測システムの開発を目的とし、実課題の検証としては、極限環境である南極に生息する微生物類や、病原体を媒介する節足動物を具体例に、環境適応や共進化の過程の解明を目指す。

3. 研究の方法

(1) 水平伝播遺伝子検出とその由来の系統推定ワークフロー

開発したワークフローの概要を図 1 に示す。本ワークフローでは、2 種類の大規模 BLSOM、全原核生物と各門の属レベルを対象とした BLSOM を用いて、HGT 検出と候補のその由来の系統推定を行う (以降、全原核生物を対象とした BLSOM を Prokaryotes-BLSOM、各門の属レベルの BLSOM を Genus-BLSOM とする)。原核生物および属を対象とした BLSOM は、少なくとも 10kb の配列が DDBJ / ENA / GenBank 上に登録されていた原核生物 3,157 種を対象に、各々の種ごとに、5000 塩基ごとに断片化した配列 (計 3,500,000 配列) の縮退 4 連続塩基組成を基に学習を行った。ここで、DNA データベースには 2 本鎖 DNA の片方の配列が登録されており、2 本鎖配列の選択に関する自由度に起因する影響を除くため、

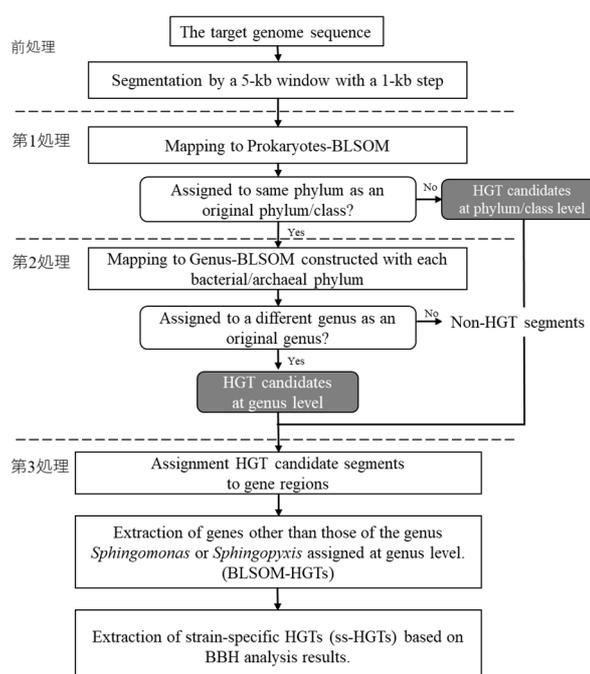


図1. BLSOMによる水平伝播遺伝子検出とその由来の系統推定ワークフロー

相補的な連続塩基（例えば AAAA と TTTT）を同一のものとみなすことを、「縮退」と定義している。

前処理として、予測対象の微生物ゲノムに対し、5kb 配列（ステップサイズ 1kb）に断片化した配列を入力データとして使用する（図 1 の前処理）。これらの入力データに対し、Prokaryotes-BLSOM マップ上の最小のユークリッド距離を持つ格子点を特定した。特定された格子点とその周辺の格子点に分類されていた既知生物の系統情報が特定の Phylum/Class が 60%以上を占める場合に、その候補由来 Phylum/Class とした（図 1 の第 1 処理）。ここで、対象ゲノムが属する Phylum/Class と異なる場合は、Phylum/Class レベルが異なる HGT 候補とし、同じ由来であった場合は、次の処理で属レベルでの系統推定を行う。対象ゲノムが属する Phylum/Class にて作成した Genus-BLSOM マップを対象に、最小のユークリッド距離を持つ格子点を特定し、特定された格子点とその周辺の格子点に分類されていた既知生物の系統情報が特定の Genus が 60%以上を占める場合に、その断片配列の候補由来 Genus とした（図 1 の第 2 処理）。ここで、対象ゲノムと異なる Genus 由来であった場合を、属レベルでの HGT 候補配列とした。

最後に、得られた HGT 候補断片配列上に含まれる遺伝子領域を抽出し、これらを最終的には水平伝播遺伝子候補とする（図 1 の第 3 処理）（この HGT 候補を、以降 BLSOM-HGTs とする）。また、近縁種間比較における水平伝播候補遺伝子の絞り込み方法として、対象ゲノムを含む同属異種ゲノムを対象に、配列相同性検索結果から、双方向ベストヒット（BBH）関係を求め、ベストヒットの関係が、自分自身にのみ得られた遺伝子と、BBH が成立できなかった遺伝子を、対象ゲノム特異的 HGT 候補として検出した（この HGT 候補遺伝子を、以降、ss-HGT とする）。

4. 研究成果

(1) 南極由来 *Sphingomonas* 属 2 株を対象にした HGT 検出

南極由来株である *Sphingomonas* sp. HMP6 と HMP9 の 2 株（以下、南極株）と、南極大陸以外の大陸で単離されて完全ゲノムが解読されている 5 株（以下、大陸株）を対象に、BLSOM を用いて断片配列に対する系統推定を行った（図 2）。系統推定結果として、まず、最下層である Genus レベルにおいて、大陸株では、平均同定割合は、82.8%であ

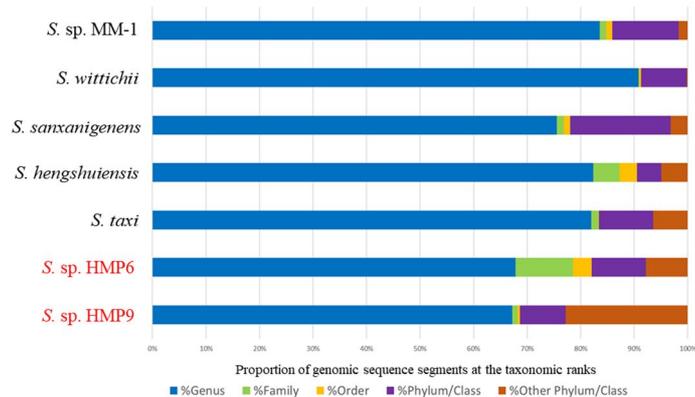


図 2. 大陸株（黒）と南極株（赤）の断片配列に対する HGT 検出と系統推定結果。各細菌ごとに予測された系統の割合を示す。系統と色の対応は、グラフ下に示す。

った。一方、南極株において、南極由来 HMP6 株の同定割合は 67.8%，南極由来 HMP9 株は 67.2%であった。南極由来 HMP6 と HMP9 株は、Genus レベルで本来の *Sphingomonas* 属 (genus) と同定される割合が、大陸株と比べて少ない傾向にあった。次に、Family (科) レベルにて、大陸株で平均は 84.6%であったのに対し、南極由来 HMP6 株では 78.5%で、南極由来 HMP9 株では 68.5%であった。最後に、Bacteria の最上位階層である Phylum/Class レベルにて、同じ Phylum/Class である *Alphaproteobacteria* と同定されたゲノム配列断片数の割合を確認したところ、大陸株で平均 96.7%であったのに対し、南極由来 HMP6 株では 92.2%と同様に *Alphaproteobacteria* への同定割合が高かったのに対し、南極由来 HMP9 株では 77.1%と南極由来 HMP6 株や大陸株と比べて低く、他の Phylum 由来への予測割合が 22.8%と異なる傾向を示した。

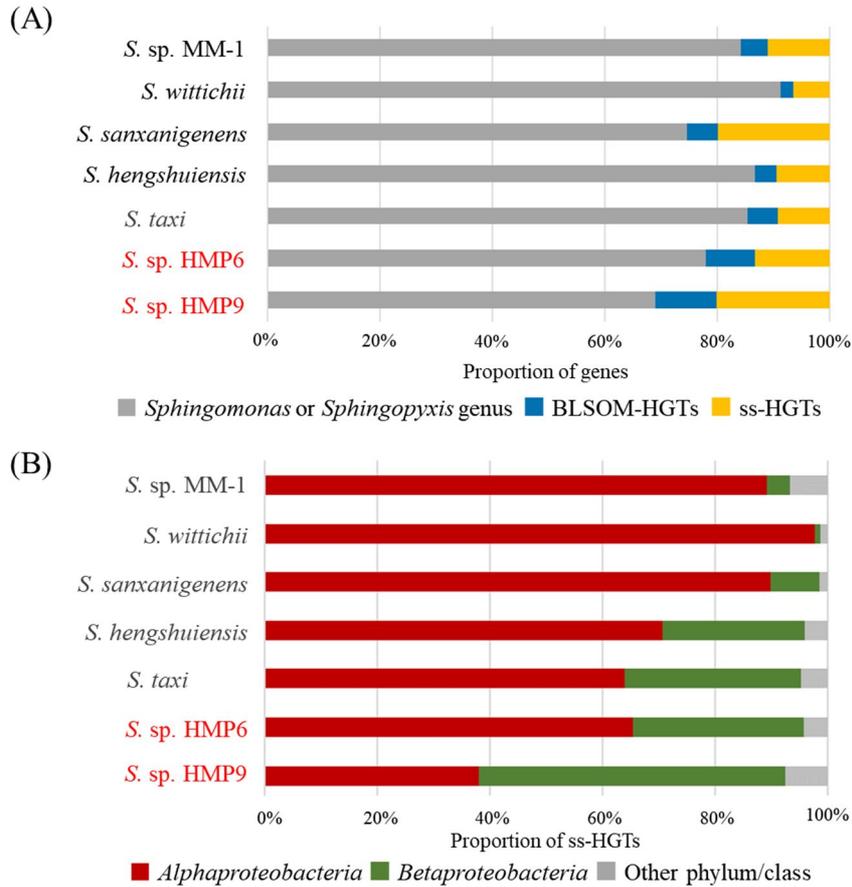


図3. 大陸株（黒）と南極株（赤）の遺伝子に対するHGT検出と系統推定結果。
 (A)全遺伝子に対する系統推定結果。(B) ss-HGTsに対する系統推定結果。
 各細菌ごとに予測された系統の割合を示す。系統と色の対応は、グラフ下に示す。

次に、ゲノム断片配列に対する系統推定結果を基に、各遺伝子に対する系統同定を行った（図3（A））。ここで、BLSOM解析によって得られた各細菌での水平伝播候補遺伝子を、BLSOM-HGTsとする。BLSOM-HGTsの割合は、大陸株で平均は15.6%であった。一方、南極由来HMP6株は22.0%で、南極由来HMP9株は31.0%で、ゲノム断片配列に対する同定結果と同様に、南極株のほうが水平伝播遺伝子の保有率が高かった。南極環境適応に対する水平伝播候補遺伝子の関与を明らかにするため、BLSOM-HGTsから株特異的水平伝播候補遺伝子（strain specific HGT candidates, ss-HGTs）の抽出を試みた。ここで、ss-HGTsは、近縁株間でのBBH解析において、自身としかBBHが成立しなかった遺伝子とBBHが成り立たなかった遺伝子と定義した。南極株において、ss-HGTsのBLSOM-HGTsに対する割合は約63%と高く、その多くが属していた。ss-HGTsの起源生物系統の詳細をPhylum/Classレベルで調べた結果を図3（B）に示す。大陸株と南極由来HMP6株はゲノム断片配列での系統推定結果と同様に、*Sphingomonas*属細菌が属するAlphaproteobacteria網細菌由来への同定割合が一番多かった。次は、Betaproteobacteria網細菌由来であった。一方、南極由来HMP9株では大陸株と南極由来HMP6株の傾向とは異なり、Betaproteobacteria網細菌由来の割合が一番高く、全ss-HGTsの半数以上を占めていた。系統的に近縁な*S. taxi*や南極由来HMP6株に比べ、より遠縁の生物群から、多くの水平伝播候補遺伝子を獲得していたことが示唆された。

南極株において系統樹解析において近縁な種と比べ、水平伝播を多く獲得していたことが示唆され、南極由来HMP6株では、それらの多くを同じFamilyやPhylum/Class由来から獲得したと考えられる。水平伝播により獲得された物質循環やエネルギー代謝の機能遺伝子がアミノ

酸組成の変化を受けたり，南極細菌が既に低温環境へ適応していた他の細菌から遺伝子を獲得したりすることで，南極大陸のような限られた遺伝資源の条件下にあっても高い環境適応能を示し，共生・共存的な生物圏の構築を可能にしたと考えられる．本手法を活用することで，網羅的に，かつ高精度に水平伝播遺伝子の由来種を予測することが可能となり，様々な環境下での生育環境適応戦略や共生関係が及ぼしてきた共進化過程の解明に向けた基盤情報の提供も可能となる．

(2) 南極由来2株に対する開発ワークフローと相同性検索による系統推定結果の比較

水平伝播候補遺伝子を検出するための BLSOM を用いた系統推定ワークフロー（本手法）の感度と精度を評価するために，我々は，相同性検索プログラムであるアミノ酸配列を対象にした BLASTP でヒットした生物系統情報との比較を行った．BLASTP では，NCBI の non-redundant

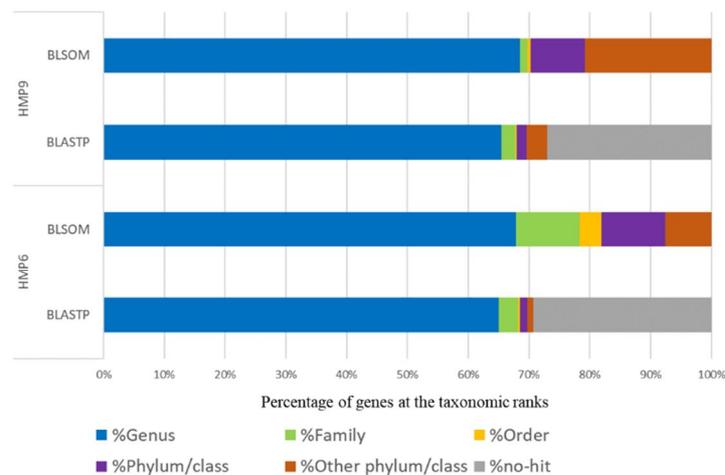


図4. 南極株の各遺伝子に対するBLSOMによる系統推定結果と相同性検索 (Blastp) による系統推定結果. 各細菌ごとに予測された系統の割合を示す. 系統と色の対応は，グラフ下に示す.

protein database (nr)を対象に，しきい値として，e-value が $1e^{-5}$ 以下でベストヒットした配列で，かつ，配列一致率と配列カバー率がともに70%以上で相同性が得られた配列の生物系統を調べた．本手法で得られた生物系統各生物系統の内訳を図4に示す．検出感度は，本手法がBLASTP と比べ高い．今回の対象とした南極株は，本解析において既知微生物との距離関係から全てのゲノム配列断片が微生物由来と同定され，それらに対し生物系統が同定されたために，高い検出感度となった．一方，相同性検索では，約30%の遺伝子が生物系統を得るための基準を満たす事ができなかった．

また，検出精度を確認するために，BLASTP で得られた生物系統との一致率の確認を，各系統レベルで行った．まず，最下層の Genus レベルで一致していた割合は，南極株は，HMP6 で66.2%，HMP9 で70.0%であった．この理由として，*Sphingomonas* 属には，*Sphingopyxis* 属など，非常に近縁な属が存在しており，BLSOM では，*Sphingomonas* 属と *Sphingopyxis* 属の識別が難しいために，BLASTP との一致割合が低くなっている．*Sphingomonas* 属と *Sphingopyxis* 属を含む場合での一致していた割合は，HMP6 で77.9%，HMP9 で71.0%と高い一致率を示していた．一方で，系統レベルを Family レベルに上げると，BLASTP との一致割合は HMP6 で81.3%，HMP9 で72.0%であった．門/クラスレベルでは，一致率は HMP6 で93.4%，HMP9 で81.1%でした．HMP9 の一致率は HMP6 よりも低い理由として，*Betaproteobacteria* 網由来細菌に由来する遺伝子を，Blastp では検出が難しく，BLSOM によってのみで割り当てられたためである．生物系統レベルに応じて，検出精度も改善していた．少なくとも，本手法は相同性検索で検出できなかった新規性の高い微生物ゲノムに対しても，網羅的に信頼性高く生物系統を同定することが可能であると言える．

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 9件/うち国際共著 1件/うちオープンアクセス 9件）

1. 著者名 Abe Takashi, Akazawa Yu, Toyoda Atsushi, Niki Hironori, Baba Tomoya	4. 巻 11
2. 論文標題 Batch-Learning Self-Organizing Map Identifies Horizontal Gene Transfer Candidates and Their Origins in Entire Genomes	5. 発行年 2020年
3. 雑誌名 Frontiers in Microbiology	6. 最初と最後の頁 1486
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fmicb.2020.01486	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 Yongjin QIU, Takashi ABE, Ryo NAKAO, Kenro SATOH, Chihiro SUGIMOTO	4. 巻 81
2. 論文標題 Viral population analysis of the taiga tick, Ixodes persulcatus, by using Batch Learning Self-Organizing Maps and BLAST search	5. 発行年 2019年
3. 雑誌名 Journal of Veterinary Medical Science	6. 最初と最後の頁 401-410
掲載論文のDOI（デジタルオブジェクト識別子） 10.1292/jvms.18-0483	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 Takashi Abe, Ryo Ikarashi, Masaya Mizoguchi, Masashi Otake, Toshimichi Ikemura	4. 巻 95
2. 論文標題 A strategy for predicting gene functions from genome and metagenome sequences on the basis of oligopeptide frequency distance	5. 発行年 2020年
3. 雑誌名 Genes & Genetic Systems	6. 最初と最後の頁 11-19
掲載論文のDOI（デジタルオブジェクト識別子） 10.1266/ggs.19-00041	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 Amano Nanako, Yamamuro Ayaka, Miyahara Morio, Kouzuma Atsushi, Abe Takashi, Watanabe Kazuya	4. 巻 68
2. 論文標題 Methylomusa anaerophila gen. nov., sp. nov., an anaerobic methanol-utilizing bacterium isolated from a microbial fuel cell	5. 発行年 2018年
3. 雑誌名 International Journal of Systematic and Evolutionary Microbiology	6. 最初と最後の頁 1118~1122
掲載論文のDOI（デジタルオブジェクト識別子） 10.1099/ijsem.0.002635	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 Torii Shiho, Orba Yasuko, Hang ' ombe Bernard M., Mweene Aaron S., Wada Yuji, Anindita Paulina D., Phongphaew Wallaya, Qiu Yongjin, Kajihara Masahiro, Mori-Kajihara Akina, Eto Yoshiki, Harima Hayato, Sasaki Michihito, Carr Michael, Hall William W., Eshita Yuki, Abe Takashi, Sawa Hirofumi	4. 巻 250
2. 論文標題 Discovery of Mwinilunga alphavirus: A novel alphavirus in Culex mosquitoes in Zambia	5. 発行年 2018年
3. 雑誌名 Virus Research	6. 最初と最後の頁 31 ~ 36
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.virusres.2018.04.005	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 Miyachi Tomoko, Kouzuma Atsushi, Abe Takashi, Watanabe Kazuya	4. 巻 7
2. 論文標題 Complete Genome Sequence of Acidithiobacillus ferridurans JCM 18981	5. 発行年 2018年
3. 雑誌名 Microbiology Resource Announcements	6. 最初と最後の頁 e01028018
掲載論文のDOI (デジタルオブジェクト識別子) 10.1128/MRA.01028-18	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 Yuki Iwasaki, Takashi Abe, Kenosuke Wada, Yoshiko Wada and Toshimichi Ikemura.	4. 巻 92
2. 論文標題 An artificial intelligence approach fit for tRNA gene studies in the era of big sequence data.	5. 発行年 2017年
3. 雑誌名 Genes & Genetic Systems	6. 最初と最後の頁 43-54
掲載論文のDOI (デジタルオブジェクト識別子) 10.1266/ggs.16-00068	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 Atsushi Kouzuma, Maho Tsutsumi, Shun ' ichi Ishii, Yoshiyuki Ueno, Takashi Abe and Kazuya Watanabe.	4. 巻 7
2. 論文標題 Non-autotrophic methanogens dominate in anaerobic digesters	5. 発行年 2017年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 1510
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-017-01752-x	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 Daniel J. Shea, Motoki Shimizu, Namiko Nishida, Eigo Fukai, Takashi Abe, Ryo Fujimoto and Keiichi Okazaki.	4. 巻 18
2. 論文標題 IntroMap: a signal analysis based method for the detection of genomic introgressions.	5. 発行年 2017年
3. 雑誌名 BMC Genetics	6. 最初と最後の頁 101
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12863-017-0568-5	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6件 (うち招待講演 6件 / うち国際学会 0件)

1. 発表者名 Takashi Abe, Yu Akazawa, Tomoya Baba, Toshimichi Ikemura.
2. 発表標題 An AI strategy for detecting HGT relies on nucleotide compositional features.
3. 学会等名 第42回日本分子生物学会年会 (招待講演)
4. 発表年 2019年

1. 発表者名 阿部貴志
2. 発表標題 ウイルスゲノムの連続塩基組成を活用した効率的な知識発見
3. 学会等名 進化学会第20回大会 (招待講演)
4. 発表年 2018年

1. 発表者名 阿部貴志
2. 発表標題 ゲノムデータベース、ゲノムアノテーション
3. 学会等名 統合データベース講習会AJACS越後 (招待講演)
4. 発表年 2018年

1. 発表者名 阿部貴志
2. 発表標題 微生物ゲノムとメタゲノムからの効率的な知識発見に向けて
3. 学会等名 木村資生記念 第2回進化学セミナー（招待講演）
4. 発表年 2018年

1. 発表者名 阿部貴志
2. 発表標題 シニア研究者の専門知識を活用したtRNA遺伝子データベースtRNADB-CEの運用.
3. 学会等名 日本遺伝学会第89回大会（招待講演）
4. 発表年 2017年

1. 発表者名 阿部貴志
2. 発表標題 ゲノムデータからの効率的な知識発見に向けた一括学習型自己組織化マップの活用.
3. 学会等名 第27回日本数理生物学会年会（招待講演）
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

水平伝播候補検出ソフトウェア http://bioinfo.ie.niigata-u.ac.jp/?PEMS4HGT 新潟大学工学部知能情報システムプログラムバイオインフォマティクス研究室 http://bioinfo.ie.niigata-u.ac.jp/

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 分 担 者	池村 淑道 (Ikemura Toshimichi) (50025475)	長浜バイオ大学・バイオサイエンス学部・客員教授 (34204)	

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