

生物多様性を俯瞰するための大規模ゲノム情報基盤の整備

課題責任者
阿部 貴志

新潟大学工学部

著者

阿部 貴志*¹, 池村 淑道*²

*¹新潟大学工学部, *²長浜バイオ大学

次世代シーケンサーの登場以降、ゲノム解読プロジェクトに拍車がかかり、ゲノム配列データの増加は想像を遥かに超える。大量のゲノム配列データを包括的、かつ俯瞰的に把握する強力な手法が必要となった。我々はこれまで、連続塩基組成に着目するだけでゲノム配列断片を高精度にクラスタリング（自己組織化）し、自己組織化に寄与した要素を可視化可能な、一括学習型自己組織化マップ（BLSOM）を開発してきた。この教師無しで説明可能型の AI（unsupervised and explainable AI）を様々なゲノム科学分野に応用することで、新規性の高い知識発見を可能にしている。BLSOM は並列計算に適しており、大規模解析が可能であるが、データ量の増加に伴い、年度ごとの更新には膨大な計算資源と計算時間が必要となっている。我々は、ES を活用した大規模解析結果の応用として、メタゲノム配列データからの病原性ウイルス配列の検出とその生物系統推定法の開発を行った。社会的な重要性の高い課題に関して、超大規模ゲノム配列データからの効率的な知識発見を可能にするための、新規性の高い情報基盤を提供している。

キーワード：一括学習型自己組織化マップ, 連続塩基組成, 環境メタゲノム, 病原性ウイルス, 説明可能型 AI

1. はじめに

次世代シーケンサーに代表されるゲノム解読技術のハイスループット化に伴い、多種多様な生物種に対するゲノム解読が行われている。大量に蓄積している広範囲の生物種のゲノム塩基配列から、生物種間の類似関係に基づき、生物種ごとの特徴を明らかにすることは、遺伝学や進化生物学を含むゲノム科学の重要課題である。

ゲノム塩基配列の解読が困難であった時期には、実験で測定可能な GC% が各生物種ゲノムや、さらにはゲノムの内部構造を特徴付ける基本的な量として用いられてきた。多くのゲノムが解読されている現在では、同じ GC% を持つゲノムが多数存在し、GC% のみでは特徴を理解するのは困難である。一方、塩基配列を文章のように扱い、単語の出現頻度解析 (Word Count) を行うことで、ゲノム配列に潜む多様な情報を効率的に抽出できる。ここで単語とは、2 連・3 連・4 連塩基のような連続塩基 (オリゴヌクレオチド) を意味する。同一の GC% を持つ生物種でも、2 連塩基について同一な出現頻度特性を持つ生物種は少なく、連続塩基が 3 連や 4 連と長くなるにつれ、同一の頻度特性を持つ生物種の可能性は極端に小さく

なる。この連続塩基頻度解析を、DDBJ/ENA/GenBank に収録されているゲノム配列の全体を対象にして、大規模な解析を行うことで、新規視点での知識発見を可能とした。

我々は、広範な生物種に由来する超大量ゲノム配列を対象に、ゲノム配列の連続塩基塩基の頻度に着目することで、生物種固有の特徴を俯瞰的に把握可能とする一括学習型自己組織化マップ (Batch-Learning Self-Organizing Map, BLSOM) を開発した [1-3]。BLSOM は生物種の情報を計算の途中で一切与えずに、連続塩基の出現頻度の類似性のみで、生物種ごとに高精度に分離 (自己組織化) する強力なクラスタリング能を持ち、その結果を容易に可視化できる。さらに、並列計算に適したアルゴリズムになっており、地球シミュレータなどの高性能計算機を用いた超大規模解析をも、世界に先駆けて可能とした [4]。ゲノム配列上には、断片配列であっても生物種を特徴づけるサイン (genome signature) が内在しており、BLSOM がそのゲノム配列の個性を見分けている。これまで、比較ゲノムによるウイルス種固有な特徴の解明、連続塩基組成に基づく生物種固有なシグナル配列の探索法、環境中の微生物叢より取得されたメタ

ゲノム配列に対する系統推定法、たんぱく質アミノ酸配列の連続アミノ酸組成に着目したたんぱく質機能推定法などへの様々なゲノム配列解析への応用を行ってきた [5-8]。

本研究報告では、開発した環境メタゲノム配列中に混在するウイルス由来メタゲノム配列に着目し、ウイルス由来メタゲノム配列の検出法、ならびに、ウイルス由来メタゲノム配列に対する系統推計法とその応用解析事例について紹介する。

2. 方法

2.1 一括学習型自己組織化マップ (Batch-Learning Self-Organizing Map, BLSOM)

コホネン博士が開発した自己組織化マップ (Self-Organizing Map, SOM) は大量で複雑な情報について、似た情報を自ずと集める (自己組織化する) ことを計算機上で実現している [9]。工学・経済学・言語学のような大量で複雑な情報を解析する分野で普及してきたが、ゲノム塩基配列の解析においては我々のグループが先導的に技術開発を進めてきた。従来型のコホネン SOM の場合、大量データの解析には長い計算時間を必要とし、出来上がった地図がデータの入力順に依存する問題があった。我々は、従来型 SOM の長所を生かしながら、再現性のある分類結果を得る形式にアルゴリズムを変更した「一括学習型自己組織化マップ (BLSOM)」を、ESI の時期より開発してきた [1, 2]。大量データに対する大規模な並列処理が可能となり、大量データ解析に適したアルゴリズムとなった [4-5]。

3. 結果と考察

3.1 BLSOM を用いたメタゲノム配列群からのウイルス由来メタゲノム配列検出ワークフローについて

本ワークフローでは、国際塩基配列データベースに格納されている全生物ゲノム配列データを対象とした BLSOM 結果 (Kingdom-BLSOM) と全ウイルスゲノム配列データを大勝した BLSOM 結果 (Virus group-BLSOM) を用いる。Virus group-BLSOM は、塩基配列長が 1kb 以上の配列データを対象に、1kb に断片化した 602, 951 件、97 種類の Family (科) のゲノム断片配列データを対象に BLSOM 解析を行った (図 1)。

300 塩基以上のメタゲノム配列を対象に、Kingdom-BLSOM ヘマッピングを行ない、マップされた格子点とその近傍に分類されていた既知生物種情報を用いて、生物ドメイン (真核生物, 原核生物, ウイルス, ミトコンドリア, 葉緑体) を推定する。ここで、ウイルス由来と推定されたメタゲノム配列を用いて、Virus group-BLSOM へのマッピングを行い、ウイルスの Family (科) レベルを推定する。

このように生物ドメインから階層的に生物系統を絞り込んでゆくことで、より詳細な系統推定が可能である。さらに、各 BLSOM ヘマッピングを行った際に、生物系統を推定できない場合もあるが、どの系統レベルかを知ることができるため、新規性の高いウイルス由来メタゲノム配列の効率的な検出ができる。

本ワークフローの検出精度を検証するために、国際塩基配列データベースに登録されたウイルスとウイルス以外 (真核生物と原核生物) のテスト用配列データセットの作成を行い、検証を行った。テスト用配列データセットは、検証を行うゲノム配列長に応じて、300~1,000 bp (データセット A), 500~1,000 bp (データセット B), 750~1,000 bp (データセット C) の 3 種類のデータセットを作成した。データセット B、および C では、約 93% 以上のゲノム断片配列が本来の生物ドメイン (ウイルス、もしくは、ウイルス以外) として検出され、ウイルス由来ゲノム配列断片データの約 80% が本来の Family (科) レベルと一致していた (表 1)。本開発ワークフローは、500bp 以上の配列長のゲノム配列断片データに対し、高精度なウイルス由来配列の検出とその生物系統予測が可能であると言える。

3.2 ダニ由来メタゲノム配列データに対するウイルスゲノム検出

我々は、北海道大学人獣共通感染症リサーチセンターの杉本千尋先生らのグループと共同で、マダニ (*Ixodes persulcatus*) が保有するウイルス叢の解明に向けたメタゲノム解析を行い、BLSOM にてウイルス由来メタゲノム配列の検出を行った [10]。メタゲノム配列データの解読後、BLSOM で解析可能な 300bp 以上の配列として、マダニの雌由来 507 配列、雄由来 324 配列が得られた。これらを対象に開発したワークフローにて、ウイルス由来メタゲノム配列の検出を行ったところ、雌雄ともに、半数以上がウイルス由来であった。また、約 2% が BLSOM

表 1. 既知ウイルスゲノムデータを対象とした予測精度

Dataset	Range of sequence lengths (bp)	Virus detection ratio (%) at Kingdom level	Coincidence ratio (%) of viral families
A	300–1,000	66.4	79.9
B	500–1,000	93.3	79.9
C	750–1,000	95.9	81.8

A, B, Cは作成した検証用データセットの名前である。



図 1. 国際塩基配列データベースに収録されているウイルスゲノム配列データを対象とした BLSOM マップ。図中、単一のウイルス科が分類されている格子点にはそのウイルス科特注の色を着色し、複数のウイルス科が分類されている格子点は黒で着色している。

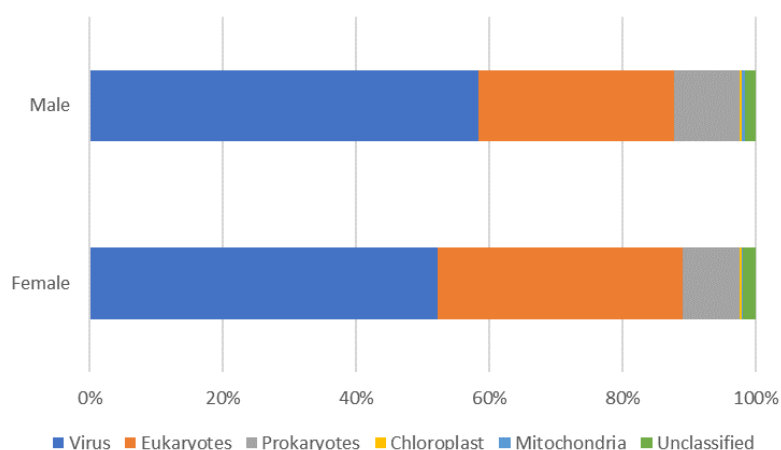


図 2. マタニメタゲノム配列データに対する生物系統推定結果の割合

にて由来を推定することができず、これらは新規性の高い生物種由来である可能性が高いと考えられる。検出されたウイルス由来メタゲノム配列に対し、Family (科) レベルでの生物系統推定を行ったところ、*Ascoviridae*, *Baculoviridae*, *Glosteroviridae*, *Iridoviridae*,

Polydnaviridae, *Poxviridae* といった昆虫に感染するウイルス科が多く検出されていた。これらは、人間または動物の病原体やダニなどの媒介性節足動物との共生関係が報告されており、これらのウイルス科に着目することで農業、および医学分野での疫病制御などへの応用

も期待できる。

従来研究で行われる配列相同性検索による生物系統推定では、推定される生物系統に偏りが生じることもあり、開発した生物系統ワークフローも活用することで、より多様なウイルス叢の把握も可能となる。

4. まとめ

我々が開発してきたゲノム配列解析手法である一括学習型自己組織化マップ (BLSOM) を要素技術として、ES を中心とした HPC を用いて、様々なメタゲノム解析により取得された大量メタゲノム配列データを対象に、微生物生態系理解のための生物系統推定、ゲノム毎の再構築を行い、加えて有用遺伝子探索のためのタンパク質機能推定手法等を開発してきた。開発した情報解析システムを活用して全地球レベルでの多様な環境における微生物生態系と生物浄化能の全体像が把握することで、各々の環境に応じた環境の保全や修復を目的とした生物浄化能の最適化が可能となり、グリーン・イノベーションにおける生物浄化能の利用促進にもつながる。地球環境科学への更なる貢献が期待できる。

次世代シーケンサーの活用に伴い、ゲノムならびにメタゲノム解析は情報爆発の時代を迎えた。ゲノムビッグデータに対応するためには、より高速化した解析手法の開発が求められている。BLSOM の可視化や分離能などの特長を生かして、爆発的なゲノム配列データの増加に対応できる新規解析手法を開発しており、ゲノムビッグデータからの効率的なデータマイニング手法として、自然科学分野のみならず、産業界や医学分野など広い分野での更なる応用を目指した研究開発を行っていきたい。

謝辞

本研究は、JSPS 科研費 17K00401 の助成を受けたものです。本研究成果は、地球シミュレータを主に用いて得られたものです。

文献

1. S. Kanaya, M. Kinouchi, T. Abe, Y. Kudo, Y. Yamada, T. Nishi, H. Mori, and T. Ikemura, "Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM) - characterization of horizontally transferred genes with emphasis on the E. coli O157 genome," *Gene*, vol. 276, pp. 89-99, 2001.
2. T. Abe, S. Kanaya, M. Kinouchi, Y. Ichiba, T. Kozuki, and T. Ikemura, "Informatics for unveiling hidden genome signatures," *Genome Research*, vol. 13, no. 4, pp. 693-702, 2003.
3. T. Abe, H. Sugawara, M. Kinouchi, S. Kanaya, and T. Ikemura, "Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples," *DNA Research*, vol. 12, no. 5, pp. 281-290, 2005.
4. T. Abe, H. Sugawara, S. Kanaya, and T. Ikemura, "Sequences from almost all prokaryotic, eukaryotic, and viral genomes available could be classified according to genomes on a large-scale Self-Organizing Map constructed with the Earth Simulator," *Journal of the Earth Simulator*, vol. 6, pp. 7-23, 2006.
5. Y. Iwasaki, T. Abe, K. Wada, Y. Wada, and T. Ikemura. A Novel Bioinformatics Strategy to Analyze Microbial Big Sequence Data for Efficient Knowledge Discovery: Batch-Learning Self-Organizing Map (BLSOM). *Microorganisms*, Vol. 1, pp. 137-157, 2013.
6. Y. Iwasaki, T. Abe, Y. Wada, K. Wada and T. Ikemura. Nobel bioinformatics strategies for prediction of directional sequence changes in influenza virus genomes and for surveillance of potentially hazardous strains. *BMC Infectious Diseases*, 13:386, doi:10.1186/1471-2334-13-386, 2013.
7. R. Nakao, T. Abe, A.M. Nijhof, S. Yamamoto, F. Jongejan, T. Ikemura, and C. Sugimoto. A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks. *ISME Journal*, Vol. 7, pp. 1003-1015, 2013.
8. T. Abe, S. Kanaya, H. Uehara, and T. Ikemura, "A Novel Bioinformatics Strategy for Function Prediction of Poorly-Characterized Protein Genes Obtained from Metagenome Analyses," *DNA Research*, vol. 16, no. 5, pp. 287-297, 2009.
9. T. Kohonen, "The self-organizing map," *Proceeding of IEEE*, vol. 78, pp. 1464-1480, 1990.
10. Y. Qiu, T. Abe, R. Nakao, K. Satoh and C. Sugimoto. Viral population analysis of the taiga tick, *Ixodes persulcatus*, by using Batch Learning Self-Organizing Maps and BLAST search. *Journal of Veterinary Medical Science*, Vol. 81, pp. 401-410, 2019.

A Large-Scale Batch-Learning Self-Organizing Map for Surveillance of Virus Community Structures

Project Representative

Takashi Abe Faculty of Engineering, Niigata University

Authors

Takashi Abe *¹, Toshimichi Ikemura *²

*¹ Faculty of Engineering, Niigata University, *² Nagahama Institute of Bio-Science and Technology

We have previously modified the conventional Self-Organizing Map (SOM), on the basis of batch-learning SOM, for genome and protein informatics, which makes the learning process and resulting map independent of the order of data input. BLSOM thus developed became suitable for actualizing high-performance parallel-computing and revealed species-specific characteristics of oligonucleotides (e.g., tetranucleotides) frequencies in individual genomes, permitting clustering (self-organization) of genomic fragments (e.g., 5 kb or less) according to species without species information during the calculation. Using ES, we established the alignment-free clustering method BLSOM that could analyze far more than 100,000,000 sequences simultaneously; sequence fragments from almost all prokaryotic, eukaryotic, and viral genomes currently available could be classified (self-organized) according to phylotypes on a single two-dimensional map. In this report, we here apply this large-scale BLSOM to identifying virus genome sequences in metagenomes derived from Ticks.

Keywords : Batch-Learning Self-Organizing Map, Oligonucleotide, Metagenome, Virus, AI

1. Introduction

One of the most important current task of life science is to unveil unknown basic knowledge from big data of genomic sequences accumulated in the International DNA Databanks. An unsupervised neural network algorithm, self-organizing map (SOM), is an effective tool for clustering and visualizing high-dimensional complex data on a single map, and we have modified the SOM for the genome analyses by developing a Batch-Learning SOM (BLSOM) [1,2]. We have used the BLSOM to analyze short oligonucleotide frequencies (di- to pentanucleotide frequency) in a wide range of prokaryotic and eukaryotic genomes [1-4].

When only fragmental sequences (e.g., 10 kb sequences) from mixed genomes derived from multiple organisms are given, it appears impossible to identify how many and what types of genomes are present in the collected sequences. However, we found that BLSOM could classify the sequence fragments according to phylotype without any information other than oligonucleotide frequencies. BLSOM recognized, in most sequence fragments, phylotype-specific characteristics of oligonucleotide frequencies, permitting phylotype-specific clustering (self-organization) of sequences and unveiling diagnostic oligonucleotides responsible for the

phylotype-specific clustering [3,4]. In previous studies, we optimized the BLSOM method for phylogenetic classification of genomic sequences obtained from mixed genomes [5-7].

In this report, we applied the large-scale BLSOM to searching virus genomes in metagenomic sequences derived from Ticks and predicting their phylogenetic origin.

2. Method

BLSOM for oligonucleotide compositions and that for peptide composition were conducted as described previously (Abe et al., 2003) [2] and (Abe et al., 2009) [8].

3. Results & Discussion

3.1 Workflow for detecting virus genome in metagenomic sequences and predicting their origin by using BLSOM

In this workflow, two types of large-scale BLSOMs, namely Kingdom- and Virus group-BLSOM, were constructed to identify viruses from metagenomic sequences using sequences deposited in DDBJ/EMBL/GenBank as previously described [2]. Kingdom-BLSOM was constructed with tetranucleotide frequencies in all 5-kb sequences derived from the whole-genome sequences of 111 eukaryotes, 2,813

prokaryotes, 1,728 mitochondria, 110 chloroplasts, and 31,486 viruses. Virus group-BLSOM was constructed with a

total of 602,951 1-kb sequences from 97 families.

Table 1. Estimation ratio of BLSOM-based classification using deposited viral sequence data.

Dataset	Range of sequence lengths (bp)	Virus detection ratio (%) at Kingdom level	Coincidence ratio (%) of viral families
A	300—1,000	66.4	79.9
B	500—1,000	93.3	79.9
C	750—1,000	95.9	81.8

Estimation ratios of BLSOM-based classification were calculated using three datasets (A, B, and C).



Figure 1. Phylogenetic classification of almost all known virus species. DegeTetra-BLSOM of 1 kb sequences derived from almost all virus genomic sequences (from 97 families). Lattice points that contain sequences only virus family are indicated in colors; those that include more than one category are indicated in black.

Virus group-BLSOM was shown in Figure 1.

After de novo assembly, contigs longer than 300 bp were mapped using Virus group-BLSOM. The mapping was conducted by finding the lattice point with the minimum Euclidean distance in the multidimensional space and was assigned to Virus group-BLSOM on the basis of statistical tests. To taxonomically classify the contigs that could not be assigned using Virus group-BLSOM, Kingdom-BLSOM was employed.

To validate accuracy of BLSOM analysis, test data sets were created from three datasets (A, B, and C) were prepared from the viral and non-viral (eukaryotic and prokaryotic) sequences deposited in DDBJ/EMBL/GenBank. The datasets A, B, and C contained BLAST-identified viral sequences with lengths ranging between 300-1,000 bp, 500-1,000 bp,

750-1,000 bp, respectively. When the dataset C was tested, approximately 80% of the fragments were correctly detected and approximately 95% of the fragments were correctly classified into the corresponding taxa by Kingdom-BLSOM at kingdom level. Furthermore, about 80% of these viral sequences were assigned to the corresponding taxa at the family level with accuracy (Table 1).

3.2 Application to detect virus genome in metagenomic sequences obtained from Tick metagenome sequences

In the report of this year, we focus on the study that has been conducted under the collaboration with Prof. Sugimoto's group (Division of Collaboration and Education, Research Center for Zoonosis Control, Hokkaido University) and

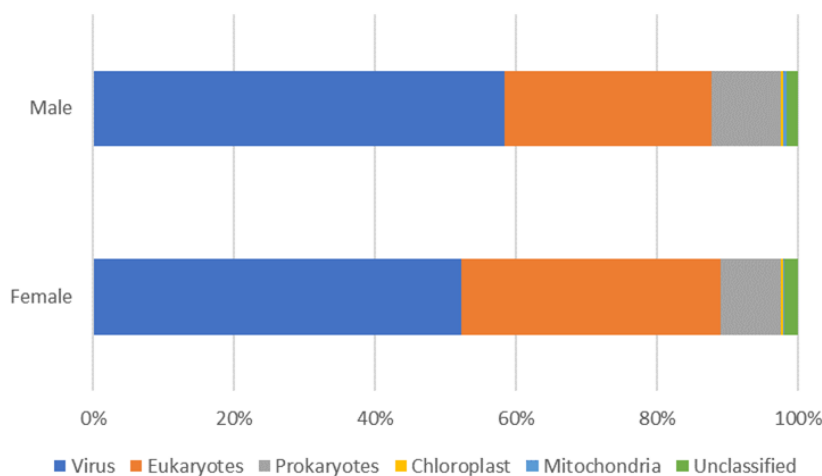


Figure 2. Proportion of kingdom taxonomy levels of the female and male samples by using Kingdom-BLSOM.

recently published by Journal of Veterinary Medical Science [9].

We focused on investigating a diversity of tick viral populations, which may contain as-yet unidentified viruses. After metagenomic sequencing, a total of 507 and 324 contigs were yielded from female and male of *Ixodes persulcatus* ticks samples, respectively, with a length of over 300 bp. BLSOM analysis of these contigs showed that about half of the contigs fell on the clusters of viruses (Figure 2). Only 2.0% (10/507) and 2.5% (8/324), from female and male samples, respectively, were not locatable in the BLSOM map (Figure 2). Sequences of double strand (ds) DNA viruses were occupied nearly 50% of viral contigs from female and male samples.

In further analysis, sequences related to several insect virus families were identified; that is, *Ascoviridae*, *Baculoviridae*, *Closteroviridae*, *Iridoviridae*, *Polydnaviridae* and *Poxviridae*. Biological interactions between human or animal pathogens, their vector arthropods and their own viruses, have been reported, and such interactions can be utilized for disease control in the fields of agriculture and medicine.

This is one of the advantages of BLSOM, when it is applied to microbiomes composed of poorly characterized and highly diversified organisms [2]. About half of the contigs were assigned to viruses using BLSOM, and only a small percentage of the contigs could not be assigned to any organisms (Figure 2), which supports that BLSOM is

theoretically advantageous in detecting and classifying previously unknown viruses over the homology-based search.

4. Conclusion

Large-scale metagenomic analyses using recently released next-generation sequencers are actively underway on a global basis, and the obtained numerous environmental sequences have been registered in the public databases. Large-scale computations using various, novel bioinformatics tools are undoubtedly needed for efficient knowledge-findings from the massive amount of sequence data.

The present BLSOM is an unsupervised algorithm that can separate most sequence fragments based only on the similarity of oligonucleotide frequencies. Unlike the conventional phylogenetic estimation methods, the BLSOM requires no orthologous sequence set or sequence alignment, and therefore, is suitable for phylogenetic estimation for novel gene sequences. It can also be used to visualize an environmental microbial community on a plane and to accurately compare it between different environments.

In summary, the oligonucleotide composition-based classification method, BLSOM, can be applied to studies of medical and veterinary important vector-arthropods, such as mosquito and fly. It should have great potential for mounting effective programs against vector-borne emerging infectious diseases.

Acknowledgement

This work was supported by Grant-in-Aid for Scientific Research (C) Grant Number 17K00401. The computation was done mainly with the Earth Simulator of Japan Agency for Marine-Earth Science and Technology.

References

1. S. Kanaya, M. Kinouchi, T. Abe, Y. Kudo, Y. Yamada, T. Nishi, H. Mori, and T. Ikemura, "Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM) - characterization of horizontally transferred genes with emphasis on the E. coli O157 genome," *Gene*, vol. 276, pp. 89-99, 2001.
2. T. Abe, S. Kanaya, M. Kinouchi, Y. Ichiba, T. Kozuki, and T. Ikemura, "Informatics for unveiling hidden genome signatures," *Genome Research*, vol. 13, no. 4, pp. 693-702, 2003.
3. T. Abe, H. Sugawara, M. Kinouchi, S. Kanaya, and T. Ikemura, "Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples," *DNA Research*, vol. 12, no. 5, pp. 281-290, 2005.
4. T. Abe, H. Sugawara, S. Kanaya, and T. Ikemura, "Sequences from almost all prokaryotic, eukaryotic, and viral genomes available could be classified according to genomes on a large-scale Self-Organizing Map constructed with the Earth Simulator," *Journal of the Earth Simulator*, vol. 6, pp. 7-23, 2006.
5. Y. Iwasaki, T. Abe, K. Wada, Y. Wada, and T. Ikemura. A Novel Bioinformatics Strategy to Analyze Microbial Big Sequence Data for Efficient Knowledge Discovery: Batch-Learning Self-Organizing Map (BLSOM). *Microorganisms*, Vol. 1, pp. 137-157, 2013.
6. Y. Iwasaki, T. Abe, Y. Wada, K. Wada and T. Ikemura. Nobel bioinformatics strategies for prediction of directional sequence changes in influenza virus genomes and for surveillance of potentially hazardous strains. *BMC Infectious Diseases*, 13:386, doi:10.1186/1471-2334-13-386, 2013.
7. R. Nakao, T. Abe, A.M. Nijhof, S. Yamamoto, F. Jongejan, T. Ikemura, and C. Sugimoto. A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks. *ISME Journal*, Vol. 7, pp. 1003-1015, 2013.
8. T. Abe, S. Kanaya, H. Uehara, and T. Ikemura, "A Novel Bioinformatics Strategy for Function Prediction of Poorly-Characterized Protein Genes Obtained from Metagenome Analyses," *DNA Research*, vol. 16, no. 5, pp. 287-297, 2009.
9. Y. Qiu, T. Abe, R. Nakao, K. Satoh and C. Sugimoto. Viral population analysis of the taiga tick, *Ixodes persulcatus*, by using Batch Learning Self-Organizing Maps and BLAST search. *Journal of Veterinary Medical Science*, Vol. 81, pp. 401-410, 2019.