

令和元年6月24日現在

機関番号：14401

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K09933

研究課題名(和文)メタゲノミック診断の個別化を目指す小規模並列解析プラットフォームの構築

研究課題名(英文) Small-scale Parallel Computing Platform for In-house Metagenomic Detection System

研究代表者

中村 昇太 (Nakamura, Shota)

大阪大学・微生物病研究所・特任准教授(常勤)

研究者番号：90432434

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)：メタゲノム解析による網羅的病原体検出法「メタゲノミック診断法」は、次世代シーケンズの簡易化や普及と共に多くの機関で行われるようになった。しかし、未だそのデータ解析から検出までの発見プロセスは、多くのノウハウや計算機資源等の障壁が存在する。本研究では、メタゲノミック診断が各医療機関で実現可能な網羅的な病原体検出用の解析プラットフォームSpark-BLASTを開発した。開発したシステムは従来使われてきたHadoopによる分散システムよりも高速な処理が可能であり、また計算ノードを追加した際の性能向上も高い線形性を示した。この解析システムを用いて実際の臨床検体データにも適用し、病原体探索を行った。

研究成果の学術的意義や社会的意義

国内では未だに小児科、救急、院内感染、臓器移植などの医療現場では感染症が疑われる不明疾患が多い。これら不明疾患において、既知・未知問わずメタゲノム解析により病原体を検出し、病態との関連性を各医療機関で個別に明らかにすることができれば今後の医療現場への貢献は計り知れない。次世代シーケンズ技術の普及が順調に進行している中、メタゲノム解析による病原体検出の一般化が達成されない理由は明らかに解析過程がボトルネックになっていることは明白である。これまで一部の大型研究施設で可能であった本方法論の個別化を達成することができれば、感染症診断の大きな進歩に繋がる可能性がある。

研究成果の概要(英文)：With the appearance of next generation sequencing (NGS) technologies, the volume of data presented to biologists has exploded. The NGS data is hard to analyze due to its sheer volume and often requires supercomputers to process in a timely manner. To face this increase, researchers have been using multi-threading solutions like MapReduce in a Hadoop-powered solution with HDFS, Hadoop distributed file system. However putting data into HDFS and then manipulating it is still very time-consuming. To address this issue, we developed Spark-BLAST, a system using the recently published in-memory cluster management software, Apache Spark. Spark-BLAST demonstrates the high efficiency of cluster computing, while proving quicker and more linear speed up than the existing solution. This offers a way to perform time-consuming similarity search with low-cost infrastructures; effectively providing the ability to handle the massive data from NGS to smaller-scale laboratories.

研究分野：感染症診断学

キーワード：メタゲノミクス 病原体検出 次世代シーケンズ 感染症診断

様式 C - 19、F - 19 - 1、Z - 19、CK - 19（共通）

1. 研究開始当初の背景

感染症は多種多様な病原体によって引き起こされるが、流行が発生した場合、その拡大を最小限に抑えるためには病原体を迅速に同定し、それに即した対策を実施することが肝要である。次世代シーケンサーによるメタゲノム解析を応用した方法（メタゲノミック診断法）では、特異的プライマーや病原体毎に異なる複雑な処理なしに、検体中の遺伝情報を網羅的に探索し病原体由来遺伝子を検出するため、既知の病原体はもちろんのこと、未知の新規病原体をも検出し得るものである（図1）。

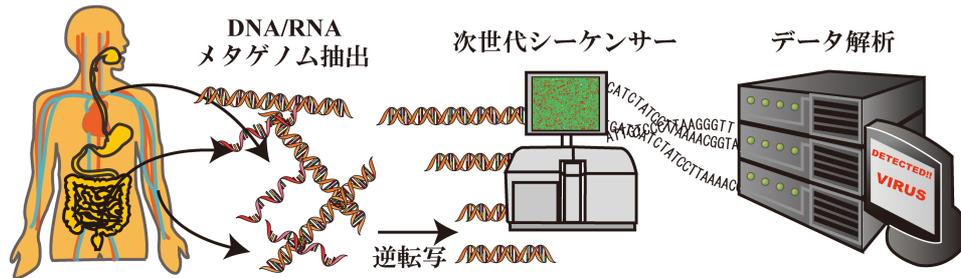


図1. メタゲノミック診断法

これまでに我々は、咽頭スワブ検体からインフルエンザウイルス、便検体からのノロウイルスの検出、また通常の検査では検出不可能であった下痢症例で病原細菌カンピロバクターの特定に成功している。これらは既知例を使用した検討であったが、最近では、我々が構築した方法論を実際の医療現場に近いところでの適用を開始している。例えば大阪大学医学系研究科の附属病院と連携し、あるがん患者の致死性敗血症例から分離され自動同定装置で肺炎桿菌が原因と判定された症例において、メタゲノム解析の結果、*Klebsiella variicola* という別の菌が原因であることを明らかにし、本菌が敗血症の病原体になりうることを示すことができた。また感染性心内膜炎例において従来法では同定できなかった候補病原体をメタゲノム解析においてのみ示すことが可能であることを明らかにした。こうした応用研究を進めてはいるが、この方法論のさらなる普及や一般化には多くの障壁がある。最近の次世代シーケンス技術の進歩と普及によって、メタゲノムデータの取得はそれほど難しい作業ではなくなったが、その膨大なシーケンス情報の解析処理に必要な計算機資源をどのように運用するのか、また病原体の検出に至るプロセスは極めて困難である。本研究課題では、このメタゲノム解析による網羅的病原体検出の方法論の個別化と一般化を目指し、小規模な医療機関でも入手可能な計算機資源を用い、手法の最適化によって高速化した解析プラットフォームの構築を目指す。最近流行しているクラウドベースの解析システムには大型計算機資源を保有したデータを集中管理する中央機関が必要であり、各医療機関はこの中央機関との情報のやりとりが必要になる。シーケンスデータがますます巨大化していく中、各医療機関から巨大データを受け取り、それを処理する計算機資源をどのように中央で運用するのかという問題が残る。我々は理想的にはデータを生み出す各医療機関で個別に解析できることが望ましいと考えている。そこでノートパソコンや安価なデスクトップPCを繋ぎあわせた小規模グリッドコンピューティングを応用し、その規模に合わせた解析手法に最適化することで現実時間の数時間内にメタゲノム解析を完了する解析プラットフォームを構築する。

2. 研究の目的

臨床検体のメタゲノム解析による網羅的病原体検出法「メタゲノミック診断法」は、次世代シーケンスの簡易化や普及と共に多くの機関で行われるようになった。しかし、未だそのデータ解析から検出までの発見プロセスは、多くのノウハウや計算機資源等の障壁が存在する。本研究では、メタゲノミック診断の個別化すなわち各医療機関で実現可能な網羅的な病原体検

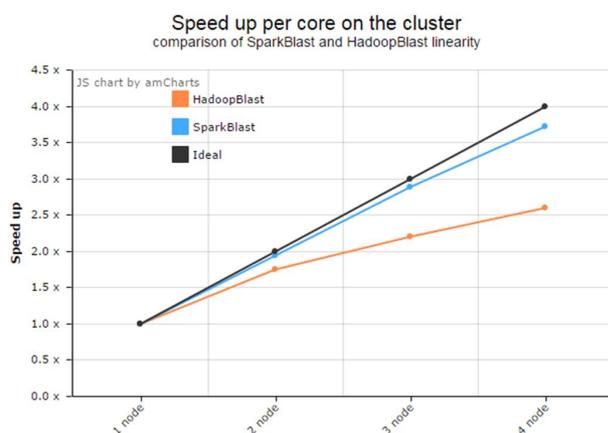
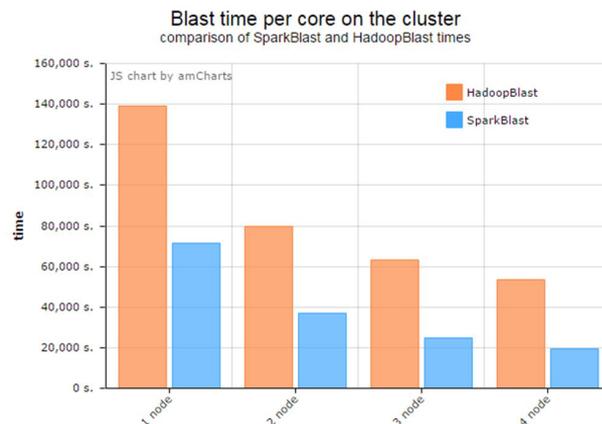
出用の解析プラットフォームの開発を目指す。研究期間内に、この解析システムの並列化を行い、小規模な医療機関でも入手可能な計算機資源を利用した臨床検体のメタゲノム解析を実現する。

3. 研究の方法

急速な次世代シーケンスの技術進歩は情報量の著しいコストダウンを達成しており、コンピュータの能力が倍増していくというムーアの法則を越える勢いで進んでいる。それに比べ、最近のコンピュータの能力の進化はムーアの法則から外れだしており、コア数の増加や省電力等の改善が為されているのみである。現在では大型計算機用のCPUの周波数と一般用とはさほど能力が変わらなくなってきておりコストも十分下がっている。こうした状況から小規模グリッドコンピューティングでも並列運用を行うことができれば、大型計算機並の能力が現実的なコストで達成可能であると考えた。そこで Apache Spark を用いて分散並行処理の基盤システムを構築する。Apache Spark は大規模なデータ処理を行うための分散処理基盤であり、従来技術である Hadoop の後継として登場した。この技術を基盤にして安価な一般用の PC を複数台接続したグリッドコンピューティングシステムを構築する。次にこのシステムで次世代シーケンスの膨大な配列情報の分散処理を行う。

4. 研究成果

構築した Apache Spark を用いた BLAST の並列処理システム SparkBLAST と、従来から開発されている Hadoop ベースの BLAST 並列処理システムの HadoopBLAST の性能評価を行った。右図は、用いた計算ノード別の 10 万本の相同性検索に要した計算時間を示している。HadoopBLAST に比べ、SparkBLAST は約 2 倍高速に並列処理することが可能であることが明らかになった。



また、左図は分散計算のノード数別の性能向上を示したプロットである。HadoopBLAST が計算ノードとして 3 ノード用いたところから性能向上の線形性が失われて、最終的に 4 ノード使用した場合で約 2.5 倍程度の性能向上しか達成されていないことに比べて、SparkBLAST では理想直線(黒)に近い線形的な性能向上を示した。小型コンピュータを多数連ねた並列処理にはこの高い線形性が必須であり、構築したシステムの有用性を明らかにした。

実際に構築した相同性検索システムが、医療機関の研究者でも実施可能かどうかをワルシャワ医科大学と共同で進めた結果、バイオインフォマティクスに不慣れな研究員でも相同性検索を行えることを確認した。ワルシャワ医科大学において継続的に使用され、病原体検出の成果を発表した(7,9,10)。また他の研究機関との微生物叢解析や病原体検出の共同研究にも用いて論文を発表した(1-6,8)。

5 . 主な発表論文等

〔雑誌論文〕(計 10 件)

- (1) Yoshizawa H, Motooka D, Matsumoto Y, Katada R, Nakamura S, Morii E, Iida T, Matsumoto H.
A case of severe soft tissue infection due to *Streptococcus tigurinus* diagnosed by necropsy in which genomic analysis was useful for clarifying its pathogenicity.
Pathol Int. 査読有 68 巻 2018 301-306 doi: 10.1111/pin.12656.
- (2) Katsimichas T, Ohtani T, Motooka D, Tsukamoto Y, Kioka H, Nakamoto K, Konishi S, Chimura M, Sengoku K, Miyawaki H, Sakaguchi T, Okumura R, Theofilis K, Iida T, Takeda K, Nakamura S, Sakata Y.
Non-Ischemic Heart Failure With Reduced Ejection Fraction Is Associated With Altered Intestinal Microbiota.
Circ J 査読有 25;82 巻 2018 1640-1650 doi: 10.1253/circj.CJ-17-1285.
- (3) Bielinska K, Radkowski M, Grochowska M, Perlejewski K, Huc T, Jaworska K, Motooka D, Nakamura S, Ufnal
High salt intake increases plasma trimethylamine N-oxide (TMAO) concentration and produces gut dysbiosis in rats.
Nutrition. 査読有 54 巻 2018 33-39 doi: 10.1016/j.nut.2018.03.004.
- (4) Uda K, Koyama-Wakai C, Shoji K, Iwase N, Motooka D, Nakamura S, Miyairi I. WU
WU polyomavirus detected in children with severe respiratory failure.
J Clin Virol. 査読有 107 巻 2018 25-28 doi: 10.1016/j.jcv.2018.08.003.
- (5) Tao C, Kinoshita N, Shoji K, Motooka D, Nakamura S, Eura R, Ueoka K, Kubota M, Ishiguro A, Miyairi I
Urinary tract infection due to anaerobic bacteria in a two-month-old infant.
J Infect Chemother 査読有 25 巻 2018 368-370 doi: 10.1016/j.jiac.2018.11.016.
- (6) Motooka D, Fujimoto K, Tanaka R, Yaguchi T, Gotoh K, Maeda Y, Furuta Y, Kurakawa T, Goto N, Yasunaga T, Narazaki M, Kumanogoh A, Horii T, Iida T, Takeda K, Nakamura S.
Fungal ITS1 Deep-Sequencing Strategies to Reconstruct the Composition of a 26-Species Community and Evaluation of the Gut Mycobiota of Healthy Japanese Individuals.
Front Microbiol 査読有 8 巻 2017 238 doi:10.3389/fmicb.2017.00238
- (7) Lipowski D, Szablowska M, Perlejewski K, Nakamura S, Bukowska-Osko I, Rzadkiewicz E, Dzieciatkowski T, Milecka A, Wenski W, Ciszek M, Debska-Slizien A, Ignacak E, Cortes KC, Pawelczyk A, Horban A, Radkowski M, Laskus T.
A Cluster of Fatal Tick-borne Encephalitis Virus Infection in Organ Transplant Setting.
J Infect Dis. 査読有 -巻 2017 - doi: 10.1093/infdis/jix040
- (8) Hagiya H, Yoshida H, Shimizu M, Motooka D, Nakamura S, Iida T, Yamamoto N, Akeda Y, Tomono K.
Herpes zoster laryngitis in a patient treated with fingolimod.
Herpes zoster laryngitis in a patient treated with fingolimod.
J Infect Chemother. 査読有 22 巻 2016 830-832 doi: 10.1016/j.jiac.2016.07.011
- (9) Bukowska-Osko I, Perlejewski K, Nakamura S, Motooka D, Stokowy T, Kosinska J, Popiel M, Ploski R, Horban A, Lipowski D, Carballo Cortes K, Pawelczyk A, Demkow U, Stepień A, Radkowski M, Laskus T.
Sensitivity of Next-Generation Sequencing Metagenomic Analysis for Detection of RNA and DNA Viruses in Cerebrospinal Fluid: The Confounding Effect of Background Contamination.
Adv Exp Med Biol. 査読有 944 巻 2017 53-62 doi: 10.1007/978-3-319-44488-8_42
- (10) Perlejewski K, Bukowska-Osko I, Nakamura S, Motooka D, Stokowy T, Ploski R, Rydzanicz M, Zakrzewska-Pniewska B, Podlecka-Pietowska A, Nojszewska M, Gogol A, Carballo Cortes K, Demkow U, Stepień A, Laskus T, Radkowski M.
Metagenomic Analysis of Cerebrospinal Fluid from Patients with Multiple Sclerosis.
Adv Exp Med Biol. 査読有 935 巻 2016 89-98 doi: 10.1007/5584_2016_25

〔学会発表〕(計0件)

〔図書〕(計2件)

(1) 中村 昇太、飯田 哲也、中屋 隆明 シーエムシー出版 メタゲノム解析技術の最前線
2017 9

(2) 元岡 大祐、中村 昇太、飯田 哲也、堀井 俊宏 羊土社
NGS アプリケーション 今すぐ始める! メタゲノム解析 実験プロトコール ヒト常在菌
叢から環境メタゲノムまでサンプル調製と解析のコツ
2016 8

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

ホームページ等

6. 研究組織

(1) 研究分担者

該当なし

(2) 研究協力者

該当なし