

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 9 日現在

機関番号：14401

研究種目：研究活動スタート支援

研究期間：2012～2013

課題番号：24890103

研究課題名(和文)第3世代シーケンサーによる感染症診断法の開発

研究課題名(英文)Development of infection diagnosis using 3rd generation sequencing

研究代表者

元岡 大祐 (Motooka, Daisuke)

大阪大学・微生物病研究所・特任研究員(常勤)

研究者番号：10636830

交付決定額(研究期間全体)：(直接経費) 2,300,000円、(間接経費) 690,000円

研究成果の概要(和文)：臨床検体からの網羅的な病原体探索法として、次(第2)世代シーケンサーを用いたメタゲノム解析法が利用されている。本手法には、生物種に依存しない統一したプロトコルを用いることが出来るため、難培養性病原体や未知病原体を探索する手法として有用である。しかし、病原体の一部のゲノム情報が得られないことが多く、感染症対策に十分な情報は得られない。そこで本研究では、新たに登場した第3世代シーケンサーを利用した感染症診断法の開発に取り組んだ。その結果、高精度かつ完全長の病原体ゲノムが容易に得られることから、感染症研究に有用であることがわかった。

研究成果の概要(英文)：Metagenomic analysis using 2nd generation sequencers is now widely used for comprehensive detection of pathogens in infectious diseases. This method with a single common protocol in clinical specimens doesn't depend on the type of microorganisms, overcoming the limitations of traditional methods. However, because this method detects only a part of genome information of pathogen, we can't get enough information to prevent infection disease. Here, I developed the method of infection diagnosis using 3rd generation sequencing. Third generation sequencing easily achieves higher degree of accuracy, less biased depth of coverage and a smaller number of contigs than other sequencing. The highly accurate complete genome analysis would develop the future genome epidemiology of pathogens.

研究分野：医歯薬学

科研費の分科・細目：基礎医学・医化学一般

キーワード：メタゲノム 次世代シーケンサー

## 1. 研究開始当初の背景

感染症は高い死因の一つであり、特に発展途上国ではマラリア、AIDS や結核などが大きな問題となっている。また救急、院内感染や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。それゆえ、感染症の病原体(細菌やウイルスなど)を迅速に検出することにより、感染症の診断、及び対策を施すことが重要である。本研究の目的は、未知のものを含むあらゆる病原体を迅速に検出する手法の構築である。

微生物の大部分は難培養性であり、分離・培養を行う手法ではごく一部の微生物しか病原体検出の対象とならない。病原体遺伝子の検査・同定法としては他に、マイクロアレイを用いた方法など、いくつかの方法が提唱されているが、これらは遺伝子情報が明らかにされた既知の病原体に対してのみ有効な方法であり、未知病原体をも同定する有力な方法は確立されていない。

これに対して、次世代シーケンサーによるメタゲノム解析では、特異的プライマーや病原体ごとに異なる処理なしに、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。

これまで我々は、第2世代(いわゆる次世代)シーケンサーを用いた病原体検出法を確立してきた(Nakamura S. et al., *PlosOne*, 2009, Nakamura S. et al., *Emerging Infectious Diseases*, 2008, Nakamura S. et al. *Exp. Biol. Med.*, 2011)。しかし、迅速性やコストなどの点において課題は残されている。また、検出できた場合でも病原体の一部のゲノム情報しか得られないことが多く、感染症対策に十分な情報は得られない。そこで本研究では、上述した問題点を解決する方法として、近年新たに登場した第3世代シーケンサーの利用を考えた。

## 2. 研究の目的

本研究は、感染性病原体の次世代シーケンサーを用いたメタゲノム解析による同定法を構築し、既知病原体のより迅速な検出、及び未知病原体の新たな検出法を確立することを目的とする。

これまでに多くの病原体検出法が開発されてきたが、感染症の拡大を最小限に抑えるためには、あらゆる病原体を統一したプロトコルで迅速に同定し、それに即した対策を実施することが肝要である。また、未知病原体の同定にとって第一選択となる有力な方法は確立されていない。

既に確立しつつある第2世代シーケンサーによる病原体検出法に加え、第3世代シーケンサーの特徴を生かすことで、より迅速安価かつ高精度での病原体同定法を確立できると考える。また、病原体の検出のみならず、迅速に病原体の全ゲノムを高精度で同定し、薬剤耐性遺伝子、病原性遺伝子やゲノム

の構造変化に関する情報を得ることは、感染症対策に大きく貢献する可能性を秘める。

## 3. 研究の方法

本研究ではまず、第3世代シーケンサーによる病原体ゲノム同定法の確立、性能評価そして第2、第3世代シーケンサーを併用した迅速な検出法を確立する。また、臨床応用のため、確立した手法をもとに様々な原因不明感染症への感染症診断法を適用し、感染症診断法の効率化と精度向上を目指す。第2世代シーケンサーとしては、Illumina 社から発売されている卓上型のシーケンサーである MiSeq を、第3世代シーケンサーとしては、Pacific Biosciences 社から発売されている PacBio RS(以下 PacBio)を使用する。

### (1)第3世代シーケンサーを用いた微生物ゲノム同定法の確立

第3世代シーケンサー(PacBio)は、第2世代シーケンサーと比較すると、迅速性や低コスト性という点で優れているということ以外に、最長 10,000 塩基を超えるロングリードが可能であるという大きな特徴がある。これは、第2世代シーケンサーが最も長くても数百塩基しか解読できないことと比較すると最も大きな特徴といえる。

この特長ゆえ、微生物の全ゲノム解析に有利であることが予想される。しかし PacBio は、発売されて間もないシーケンサーであること、また世界的に導入数が僅かであることなどから、サンプル調製法、解析方法を検討する必要がある。そこでまず、解読に適した配列長や DNA 断片化法(ネビュライザー、ソニックプローブやハイドロシェアなど)を検討し、第3世代シーケンサーによる微生物全ゲノム同定に適した条件を確立する。予備の実験としてまず、細菌ゲノムを用いる。次いで、PacBio により得られるロングリードに適した解析システムの開発も行う。特に方法論の見直し、分散化など迅速化に取り組む。これらの工程を自動的に処理する解析パイプラインの構築を目指す。

### (2)細菌ゲノム同定における第2、第3世代シーケンサーの性能比較

第2世代シーケンサー(MiSeq)からは、ショートリード(< 250 bp)しか得られないが、1,000 万本以上の大量の塩基配列を一度に解読することが出来る。その為、これまで全ゲノム同定法として広く用いられ、多数の細菌ゲノムの *DeNovo* アセンブリに使用されてきた。しかし多くの場合、5 Mbp 程度の細菌ゲノムですら、数十本のコンティグにまでに分かつなぎ合わせることが出来ておらず、十分な情報が得られているとは言い難い。一方で、第3世代シーケンサー(PacBio)は、ロングリードが得られるためゲノムのアセンブリには有利であるが、塩基解読精度は悪く、また数万本程度の塩基配列しか得られない。そ

ここで、(1)で行った PacBio を用いたゲノムアセンブリの結果を、MiSeq のデータのみや MiSeq と PacBio のデータを併せて行ったゲノムアセンブリの結果と比較検討し、病原体のゲノム同定を行う上で適した方法を決定する。

### (3)第3世代シーケンサーの感染症診断法への適用

PacBio は、上述したように MiSeq と比較するとロングリードが可能であるため、ゲノムの *DeNovo* アセンブリに有利であるが、スループットの点では劣る。そのため感染症の診断法としては、まず MiSeq を使い、臨床検体中に含まれる病原体を網羅的に探索する。分離培養可能な病原体のゲノム断片が見つかった場合には、その病原体を単離、及び PacBio による全ゲノム同定を行い、既存の株と比較する。一方、未知病原体の由来と思われる断片配列が得られた場合には、これらの断片配列を参考にプライマーを設計し、出来る限り長い配列を PCR により増幅後、PacBio でゲノムの再構築を行う。

様々な国内外の感染症研究者・医療関係者とコンタクトをとり、原因不明感染症の病原体検出を進め準備を行なっている。国内においても、小児における劇症肝炎は原因不明な例が約 30%もあることが指摘されており、生体肝移植後の再発例も多く、問題になっている。こういった原因不明の感染症疑い例に対し、上述の病原体診断法の応用に取り組む。

## 4. 研究成果

### (1)第3世代シーケンサーを用いた微生物ゲノム同定法の確立

検討実験には、食中毒の原因菌である *Vibrio parahaemolyticus* のゲノムを用いた。本菌は、約 3 Mbp と 2 Mbp の 2 本の染色体で構成される。

まず、PacBio による塩基解読に適した配列長や DNA 断片化法(ネビュライザー、ソニックプローブやハイドロシエアなど)を検討した。その結果、ハイドロシエアを用いた場合に図 1 に示すような、分散度が低い DNA 断片が得られることがわかった。また、精製方法によって得られるシーケンス結果が大きく異なり、MO BIO 社の PowerSoil キットを用いた場合に最も良質な結果が得られることが明らかとなった。

このサンプルの PacBio によるシーケンシングからは、リード数が約 32,000、平均リード長が約 3,800 bp、最長リードが 10 kbp 以上のデータを得ることが出来た(図 2)。

また、*DeNovo* アセンブリ方法を様々なソ

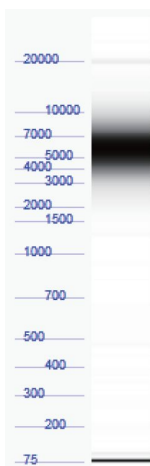


図 1 サイズ分布

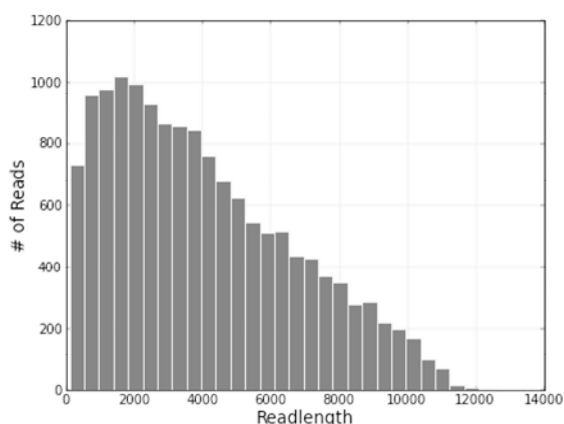


図 2 PacBio のリード長分布

フトウェアを用いて検討したところ、CeleraAssembler が最も適していることがわかった。また、ロングリード(3,000 bp 以上)の結果のみを用いてアセンブリした場合に最も効率的にアセンブリ出来ることがわかった。これらの結果を踏まえ、塩基配列データを効率よくアセンブルするパイプラインを構築した。本手法を用いた場合、*Vibrio parahaemolyticus* については、1 回のシーケンシングデータから完全長ゲノムを得ることに成功した。

### (2)細菌ゲノム同定における第 2、第 3 世代シーケンサーの性能比較

上述したように、PacBio を用いることで、細菌の完全長ゲノムを比較的容易に得ることが出来たが、PacBio のシーケンスエラーの高さゆえ、つなぎあわせたコンティグの精度を評価する必要があった。そこで同一細菌に対して、MiSeq を用いたゲノムシーケンスを行い、*DeNovo* アセンブリを行った。また、精度の高い MiSeq とリード長の長い PacBio を併用することで、より高い精度のデータが得られることを期待し、これらを組み合わせたハイブリッドアセンブリも行った。得られたデータの統計値を表 1 に示す。

表 1 MiSeq、PacBio で得られたデータの統計値

	PacBio	MiSeq
# reads	601,080	39,651,366
# bases	3,116,894,565	9,921,799,735
Coverage	x 603	x 1919
Mean length	5,815 bp	250 bp

1 回のシーケンシングにより、MiSeq からは PacBio の約 65 倍のリード数、約 3 倍の塩基数のデータを得ることが出来た。一方で、得られた平均リード長は、PacBio が圧倒的に長かった。これら得られたデータを用いて *DeNovo* アセンブリを行った結果を表 2 に示す。これらの結果から、PacBio のデータを用いた場合のみに完全長ゲノムが再構築出来ていることがわかる。MiSeq のみのデータでは、他の報告例と同様に数十本のコンティグ

にまでしかつなげ合わせることが出来ず、PacBioのデータにMiSeqのデータを組み合わせるとアセンブリに悪影響を及ぼすこともわかった。また、リファレンス配列に対する正確性についても比較したところ、ゲノムカバー率や塩基解読ミスの数という点でもPacBioのデータが最も精度が高いことがわかった(表2)。

表2 MiSeq、PacBio データのアセンブリの結果

	PacBio	PacBio /MiSeq	MiSeq
# contigs	2	7	29
Max length / Mbp	3.29	1.88	0.73
Total length / Mbp	5.17	5.17	5.09
N50 / Mbp	3.29	1.64	0.51
Genome fraction%	99.999	99.986	98.478
# mismatches	157	215	392

MiSeq では、つなぎ合わせることが出来なかった領域についても PacBio では出来ていた。各シーケンサーにより得られたコンティグをリファレンス配列にアライメントしたところ、これらの領域は主に rRNA のリピート配列が存在する領域であることがわかった(図 3)。PacBio は、これらのリピート配列を超える長いリードを解読出来るため、*DeNovo* アセンブリに有利であることがわかった。以上の結果から、細菌ゲノムの決定を行う上では、PacBio がより迅速に、より長いゲノムを再構築できることがわかった。

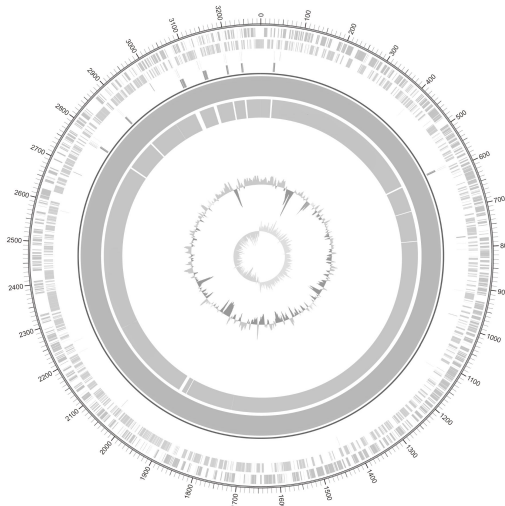


図3 染色体1のMiSeq及びPacBioのコンティグのアライメント図。外側から、forward CDS, reverse CDS, tRNA, rRNA, PacBio contig, MiSeq contigs, %GC plot, GC skew.

### (3) 第3世代シーケンサーの感染症診断法への適用

本研究中に2例の細菌感染例の臨床検体について本手法を適用することが出来た。

ケース1：感染症による流産の起因菌として同定された、ゲノムサイズが約0.7 Mbpの *Ureaplasma parvum* について、PacBio シーク

ンス及び上述手法によるゲノムの *DeNovo* アセンブリを行い、完全長のゲノム配列を得ることに成功した。

ケース2：原因菌不明の敗血症患者の臨床検体について、MiSeqを用いたメタゲノム解析を行った。その結果、敗血症の原因菌として *Klebsiella variicola* が同定できた。そこで、分離培養した本菌について、PacBioによる全ゲノム解析を行い、1回のシーケンシングにより、ほぼ完全長のドラフトゲノム(約5.5 Mbp)を得ることに成功した。

これらの症例については、今後の課題として他の株とのゲノム比較を行っていく。本研究により、臨床検体中の病原体の探索には第2世代シーケンサーを、そして全ゲノム解析には第3世代シーケンサーを用いる手法が効率的であることを明らかにした。本研究中には、既知病原体感染例や分離・培養可能例についてしか、検討を行うことは出来なかったが、上述した方法を用いれば、未知病原体や難培養性病原体に対しても、多くのゲノム情報が得ることが出来ると期待できる。本手法を用いることで、迅速かつ安価にゲノム解析を行えることから、他の病原菌や薬剤耐性菌などについても有用な手法に成りうると思える。

### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

(雑誌論文)(計4件)

Wu HN, Nakura Y, Motooka D, Nakamura S, Nishiumi F, Ishino S, Kawai Y, Tanaka T, Takeuchi M, Nakayama M, Fujita T, Yanagihara I, Complete Genome Sequence of *Ureaplasma parvum* Serovar 3 Strain SV3F4, *Genome Announc*, 2, e00256-14, 2014, 査読有, doi: 10.1128/genomeA.00256-14

Imai A, Gotoh K, Asano Y, Yamada N, Motooka D, Fukushima M, Kanzaki M, Ohtani T, Sakata Y, Nishi H, Toda K, Sawa Y, Komuro I, Horii T, Iida T, Nakamura S, Takashima S, Comprehensive metagenomic approach for detecting causative microorganisms in culture-negative infective endocarditis, *Int J Cardiol*, 172, e288-289, 2014, 査読有, doi: 10.1016/j.ijcard.2013.12.197

Seki M., Yoshida H., Gotoh K., Hamada N., Motooka D., Nakamura S., Yamamoto N., Hamaguchi S., Akeda Y., Watanabe H., Iida T., Tomono K., Severe respiratory failure due to co-infection with human metapneumovirus and *Streptococcus pneumoniae*, *Respiratory Medicine Case Reports*, 12, 13-15, 2014, 査読有, 10.1016/j.rmcr.2013.12.007

Seki M., Ohno H., Gotoh K., Motooka D.,

Nakamura S., Iida T., Miyazaki Y., Tomono K.,  
*IDCases*, **1**, 5-8, 2014, 査読有,  
doi:10.1016/j.idcr.2014.01.001

〔学会発表〕(計 5 件)

Daisuke Motooka, A Database for Rapid  
Detection of Pathogens in Infectious Diseases,  
Asian-African Research Forum on Emerging and  
Reemerging Infection 2014, 2014 年 1 月 20 日 ~  
1 月 22 日、仙台国際センター

元岡大祐、メタゲノミック診断法の高速解  
析パイプラインの構築、NGS 現場の会第三回  
研究会、2013 年 9 月 4 日 ~ 9 月 5 日、神戸国  
際会議場

宮本真理、元岡大祐、第 2・第 3 世代シー  
クエンスデータを使った CLC Genomics  
Workbench による De Novo アセンブリ、NGS  
現場の会第三回研究会、2013 年 9 月 4 日 ~ 9  
月 5 日、神戸国際会議場

朝野仁裕、元岡大祐、感染性心内膜炎にお  
けるメタゲノム解析、NGS 現場の会第三回研  
究会、2013 年 9 月 4 日 ~ 9 月 5 日、神戸国際  
会議場

後藤和義、元岡大祐、原因不明食中毒のメ  
タゲノミック診断、NGS 現場の会第三回研究  
会、2013 年 9 月 4 日 ~ 9 月 5 日、神戸国際会  
議場

〔その他〕

ホームページ等

<http://imet.gen-info.osaka-u.ac.jp/>

## 6 . 研究組織

### (1)研究代表者

元岡 大祐 (MOTOOKA, Daisuke)

大阪大学・微生物病研究所・特任研究員(常  
勤)

研究者番号：10636830