

厚生労働科学研究費補助金（新型インフルエンザ等新興・再興感染症研究事業）
バイオテロに使用される可能性のある病原体等の新規検出法と標準化に関する研究
分担研究報告書

分担研究課題： 超高速病原体ゲノム解読システムの構築と包括的な核酸迅速診断法の確立

研究分担者	黒田 誠	国立感染症研究所・病原体ゲノム解析研究センター センター長
研究協力者	関塚剛史	国立感染症研究所・病原体ゲノム解析研究センター第三室
	竹内史比古	国立感染症研究所・病原体ゲノム解析研究センター第三室
	山下明史	国立感染症研究所・病原体ゲノム解析研究センター第三室

研究要旨

未知病原体や変異病原体による新興感染症の汎発流行、そしてそれらを利用したバイオテロなどの危険性は近年社会的不安の一つとして認識されつつある。その危険性に対する確かな対処法を立案・整備する上で、バイオテロ病原体を網羅的かつ迅速に配列解読することは最も確かなアプローチの一つと考える。次世代ゲノムシーケンサー（Next-generation DNA sequencer: NGS）のパフォーマンスを用いて WHO 指定バイオテロ病原体のゲノム配列及び変異情報データベースを充実させ、有事において迅速に対応出来る体制を整えることを本研究課題は目標としている。

次世代型網羅的病原体検査法の骨格が既に構築済みであり、国立感染症研究所・病原体ゲノム解析研究センターで検査可能であるが、バイオテロ発生（および疑い事例を含む）において現場対応が迅速であればあるほど有効な対応策だと考えている。感染研・ゲノムセンターでは現場（病院・地方衛生研究所等）でも情報解析できるよう、ネットワーク経由で病原体鑑別するための情報解析パイプライン（Metagenomic Pathogen Identification pipeline for Clinical specimen: MePIC）を開発し Web 情報解析サービスを開始した。情報解析を習熟していない検査技師であっても簡単な講習会により利用できるよう利便性を考えて作成している。また、バイオテロ病原体の臨床分離株のゲノム情報を活用し、ゲノムワイドな SNPs 検索と分子系統樹作成も可能にするシステム Global core-Genome SNPs Analysis (GcoGSA)を開発し、現在、炭疽菌(*Bacillus anthracis*)のみ Web 解析サービスを開始した（関係者のみの運用予定）。順次、他カテゴリーA 病原体であるペスト菌、野兔病菌、コクシエラ、類鼻疽菌、ボツリヌス菌のゲノム情報を用いた GcoGSA を展開していく予定である。

将来的に日常の微生物検査で NGS が汎用される日が来るであろう。その際、予見しえなかった症例から NGS – MePIC – MEGAN パイプラインでカテゴリーA 病原体を検出した場合、引き続き GcoGSA にてゲノム分子疫学解析を行い、バイオテロ病原体（もしくは孤発例）の由来を推定するトレーサビリティ・追跡への情報提供になり、有事における迅速なバイオテロ対策へと貢献できると考える。

A．研究目的

未知病原体や変異病原体による新興感染症の汎発流行、そしてそれらを利用したバイオテロなどの危険性は近年社会的不安の一つとして認識されて

いる。最先端技術を駆使した次世代ゲノムシーケンサーにより、今までは数年を要した全ゲノム解読が数週間で終了することが可能となった。最先端の革新技術を応用し、効率的かつ安定的に病原体検査

システムを運用する体制を整え、WHO 指定バイオテロ病原体および未知病原体をも検査対象とする網羅的解析法を構築することを目的としている。

B. 研究方法

国立感染症研究所・病原体ゲノム解析研究センターで構築した網羅的病原体検査法の情報解析のみをパイプライン化し、インタラクティブに誰でも利用できるソフトを開発した。詳細は研究結果-3, 4を参照。

(倫理面への配慮)

該当なし

C. 研究結果

1) ネットワーク経由によるバイオテロ病原体検索のための解析システム

現在、感染研・ゲノムセンターにベンチトップ型・次世代シーケンサーと情報解析サーバーが整備されている。感染研に臨床検体が到着後、数日で解析・検査結果を報告できるようにはなっているものの、感染研に検体が送付されるまでには現場で数多くの微生物検査等が行われた後になってしまうケースが少なくない。このような現状の中、各地方自治体にも本研究課題で構築したシステムを導入して頂き、検査体制の一助となるのであれば非常に有効かつ迅速なバイオテロ対策に資するものと考えている(図1)。

本システムで特に重要なのが配列解読後の情報解析にあたる。基本、情報解析に特化した専門家が必要となるが、各自治体に人材を用意もしくは養成している余裕はない。そこで、その煩雑な解析部分をできるだけ簡便化するために、検査技師等がインタラクティブに病原体鑑別を利用できるよう、Web interface による情報解析パイプラインを用意した。

2) 配列解読から情報解析まで

ベンチトップ型・次世代シーケンサー MiSeq (Illumina)の解読リードを病院・地方衛生研究所等

でも情報解析できるよう、ネットワーク経由で病原体鑑別するための情報解析パイプライン (Metagenomic Pathogen Identification pipeline for Clinical specimen: MePIC) を構築した(図2)。情報解析を習熟していない検査技師であっても簡単な講習会により利用できるよう利便性を考えて作成した。本システム MePIC は、感染研、地方衛生研究所においても利用可能な次世代型網羅的病原体検索システムをサポートする Web 情報解析サービスであり、現在、アカデミアのみアカウント取得可能として提供している。

<https://mepic.niid.go.jp/cgi-bin/mepic2/index.cgi>

3) Metagenomic Pathogen Identification pipeline for Clinical specimen: MePIC

MiSeq 等の次世代シーケンサーの解読リードを下記の手順で必要な情報パラメーターを設定し、解析ボタンのワンクリックで解析をシームレスに行うことができる。

Target read files:

次世代シーケンサーの解読リードをアップロード

Reads trimming:

解読リードの不必要なアダプター配列の削除とクオリティーの低い塩基の削除

Screening:

bwa mapping 法によるヒト配列の削除(マウス等、他のほ乳動物のゲノム配列など、別途、レファレンス配列の指定が可能)

Reads classification:

megaBLAST によるデータベース検索。各種データベースの選択が可能。非特異的なヒットを選択しないよう、E-value 等の閾値の変更も可能

MePICで解析したMegablast解析結果は、配列アライメントを含むテキスト配列で排出される。そのテキスト配列をフリーソフト MEGAN (MEtaGenome Analyzer

<http://ab.inf.uni-tuebingen.de/software/megan/>)で生物種毎の系統分類をおこない、病原体候補を探索し

ていく。臨床検体に内在する数多くの候補生物が抽出されるため、ここでの探索には病原体と感染症の知識が多分に要求される。

4) コアゲノム SNPs を利用した菌種・菌株の類縁関係の特定 (図4)

MePIC - MEGAN による病原体検索の結果、仮に炭疽菌が候補として浮上した場合、先に使用した解読リードを用いて病原体の由来を推定するためのゲノムワイド SNPs 解析システム Global core genome SNP analysis for *Bacillus anthracis* (GcoGSA-BA) を開発した (図5)。現在、関係者のみ運用可能としている。GcoGSA-BA は、公開されている炭疽菌ゲノムと比較したゲノムワイド系統樹の生データまで作成し、適当な系統樹ビューワーで閲覧可能なデータのダウンロードを可能にする。炭疽菌は *Bacillus cereus* group に属し、セレウス菌との鑑別間違いを起さぬよう、Lethal factor (LF), Edema factor (EF), Protective antigen (PA) の炭疽菌の病原性に必須な因子の特定も可能にした。

今後、ペスト、野兔病、コクシエラ、類鼻疽、ボツリヌスと順々に構築予定である (図6)。

D/E. 考 察・結 論

これまで分担研究として、WHO 指定バイオテロ病原体の配列データベース化を進めてきた。ゲノム情報を活用することにより有効なトレーサビリティに役立てる目的である。構築データベースを有効に活用するためには、迅速な解読リード配列の取得が望ましく、そのためには多くの難題が残っている。本年度の課題として、情報解析の工程に特化して迅速性を追求したシステム改善を行った。今後、様々な工程の中でボトルネックになっている箇所を効率よく改善し、総合的なシステム化に貢献したいと考えている。NGS - MePIC - MEGAN - GcoGSA の解読・解析パイプラインを開発し、実際に運用できるところまで完了した。未だ解消されていないボトルネックが残っていることは事実だが、誰もが簡便に

利用できるシステムが先行していけば、シーケンサー等のインフラ整備が後々追いついてくるものと考えている。

F. 健康危険情報

特になし

G. 研究発表

1) 論文発表

- MePIC, Metagenomic Pathogen Identification for Clinical Specimens. Fumihiko Takeuchi, Tsuyoshi Sekizuka, Akifumi Yamashita, Yumiko Ogasawara, Katsumi Mizuta, and Makoto Kuroda. Jpn. J. Infect. Dis., 2014, 67 (1): 62-65.

2) 学会発表

なし

H. 知的財産権の出願・登録状況 (予定を含む)

該当なし

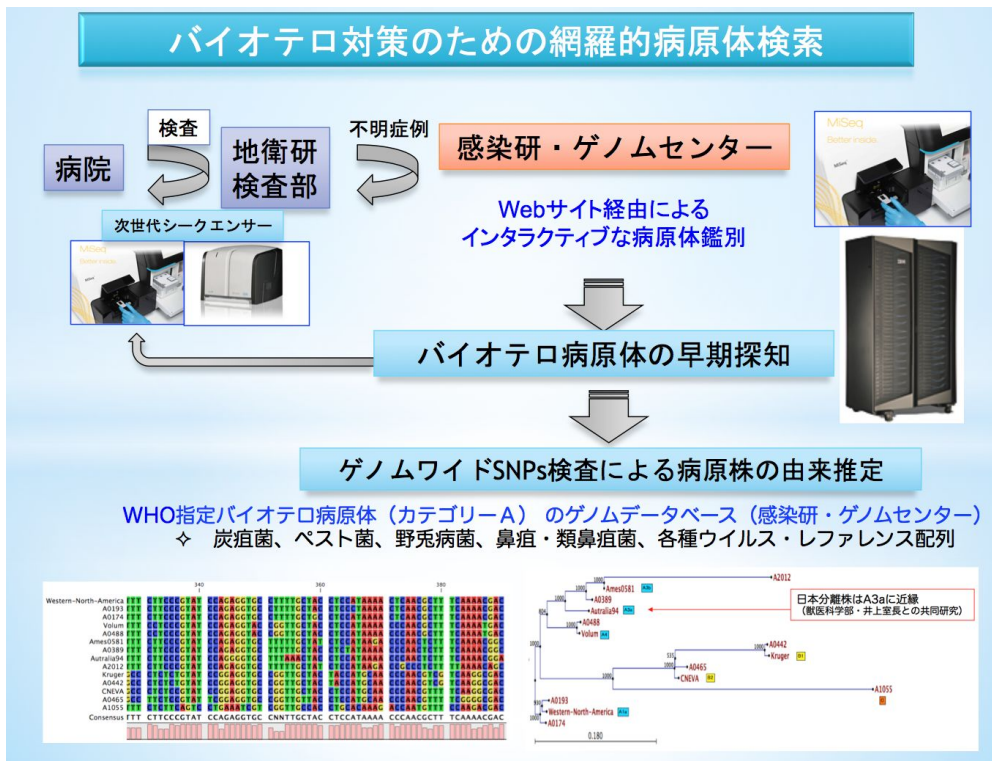


図1 ネットワーク経由によるバイオテロ病原体検索のための解析システム。現在、感染研・ゲノムセンターにベンチトップ型・次世代シーケンサーと情報解析サーバーが整備されている。地方衛生研究所等においても少しずつ整備されるようになってきた。検査技師等がインタラクティブに病原体鑑別を利用するために、Web interface による情報解析プログラムの開発と提供を試みた。

次世代型網羅的病原体検索システムをサポートする

Web情報解析サービスの提供

MePIC, Metagenomic pathogen identification for clinical specimens

<https://mepic.niid.go.jp/cgi-bin/mepic2/index.cgi>

ヒト臨床検体からの網羅配列解読(メタゲノム解読)の情報解析
 臨床検体に内在する微生物群の分類
 初心者でも使いやすい
 (生物学、感染症学の知識は欲しい)

Welcome to MePIC v2.0

Please login to start

User ID: Password:

First visit? Please [register account](#).

[to MePIC manual](#)

現在、アカデミアのみアカウント取得可能

図2 臨床検体からダイレクトに網羅配列解読したリード配列の情報解析パイプライン (Metagenomic Pathogen Identification pipeline for Clinical specimen: MePIC)。次世代シーケンサー NGS による解読リードから必要な情報のみ抽出するための Web 情報解析サービス。

図3 MePIC の操作画面の仕様。NGS リードから i) 質の悪いリード、塩基の除去、ii) 臨床検体に内在するヒト配列の削除、iii) megablast 検索による病原体検索。この i - iii までの一連の流れを一度に完結させるパイプライン。

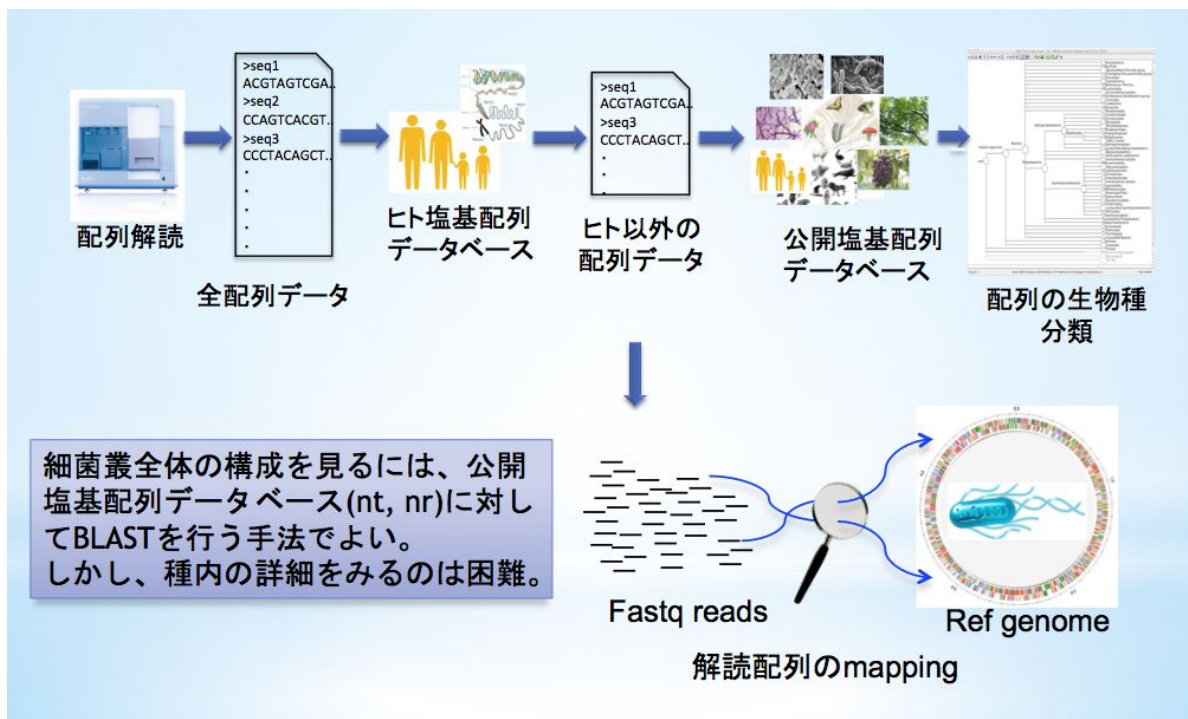


図4 ゲノム全体の SNPs (ゲノムワイド SNPs 解析) の概念図。

炭疽菌 *Bacillus anthracis* のゲノム分子疫学 (関係者のみ運用予定)

GcoGSA BA

Global core Genome SNP Analysis for *Bacillus anthracis*

Project name:

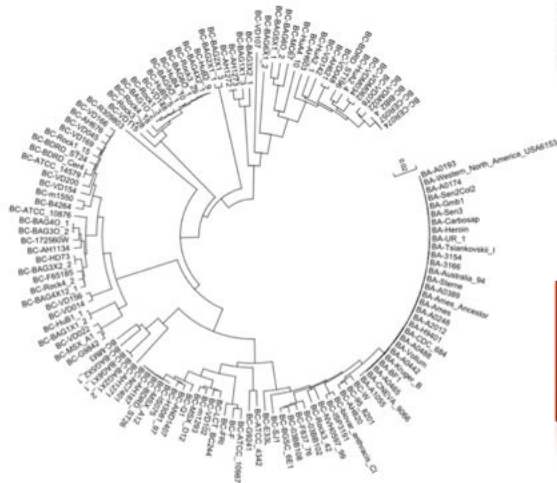
Input your sequence reads (Only .fastq.gz up to 2GBytes in total are acceptable.)

Strain Name	Read file 1	Read file 2
1	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。
2	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。
3	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。
4	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。
5	選択... ファイルが選択されていません。	選択... ファイルが選択されていません。

Send an e-mail after finished analyzing.
 Mail address:
 Change number of samples

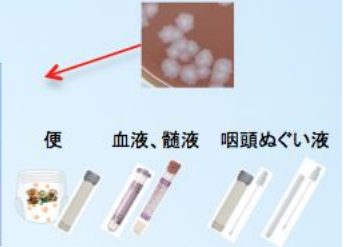
Reference genome: *Bacillus anthracis* str. 'Ames Ancestor' chromosome, complete genome. (gil50196905reflNC_007530.2)

Phylogenetic tree of the original data set.



[Traditional tree shape](#)
[Newick formatted data](#)
[MEGA formatted data](#)
[SNP allere table \(compressed 13MBytes\)](#)

LF
EF
PA



~~Gco junior~~



SNP判定のエラーが多い

Mapping status

Strain	Raw reads	After Trimming	Mapping
S6 sequence	7631281	4926893 (64.6%)	4896388 (64.2%)
S5 sequence	6850274	4814255 (70.3%)	4775247 (69.7%)

Mapped region

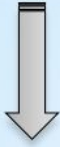
Sequence or region	Whole length	Mapped bp (%) (depth >=5)	
		S5 sequence	S6 sequence
gil50196905reflNC_007530.2l	5227419	4914301 (94.0%)	4827010 (92.3%)
gil20520075igblAE011190.1l	181677	163651 (90.1%)	161559 (88.9%)
gil50118566igblAE017335.3l	94830	84018 (88.6%)	81139 (85.6%)
lethal factor (AE011190 149357-151786)	2429	1865 (76.8%)	1767 (72.7%)
edema factor (AE011190 122608-125010)	2402	1946 (81.0%)	2187 (91.0%)
protective antigen (AE011190 143779-146073)	2294	2225 (97.0%)	2161 (94.2%)

Number of SNPs used: 636038 / 657183 (96.8%)

[Download SNP data in tabular format. \(8.6 MBytes\)](#)
[Download SNP data in fasta format. \(83.3 MBytes\)](#)
[Download phylogenetic tree in Newick format. \(5.5 KBytes\)](#)
[Download phylogenetic tree in PDF format. \(9.3 KBytes\)](#)

図5 MePIC - MEGAN による病原体検索の結果、炭疽菌が候補として浮上した場合、先に使用した解読リードを用いたゲノムワイド SNPs 解析システム Global core genome SNP analysis for *Bacillus anthracis* (GcoGSA-BA) を開発した。現在、関係者のみ運用可能としている。公開されている炭疽菌ゲノムと比較したゲノムワイド系統樹の生データまで作成し、適当な系統樹ビューワで閲覧可能なデータをダウンロード可能である。炭疽菌は *Bacillus cereus* group に属し、セレウス菌との鑑別間違いを起さぬよう、Lethal factor (LF), Edema factor (EF), Protective antigen (PA) の炭疽菌の病原因子の特定も可能にした。

illumina .fastq 生解読リード



- De novo assembly
- RNA finding
- 16S-rRNA phylogeny
- Species identification

GcoGSA : Global core Genome SNPs Analysis

GcoGSA-BA (炭疽菌、セレウス菌)

- YP(ペスト菌)
- FT(野兎病菌 Type A, B)
- CR(コクシエラ、リケッチア)
- BPM(類鼻疽菌)
- CB(ボツリヌス菌)

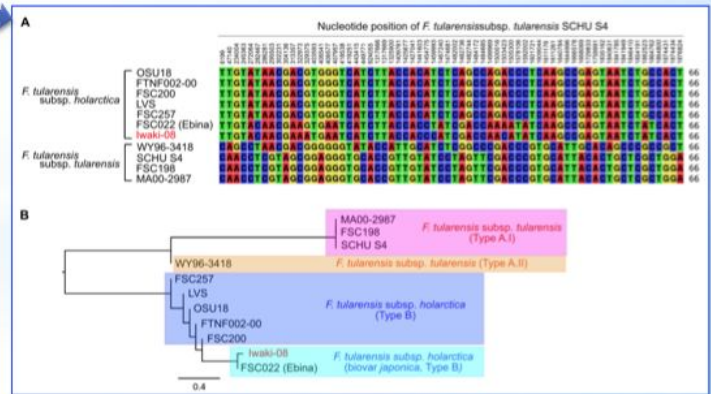
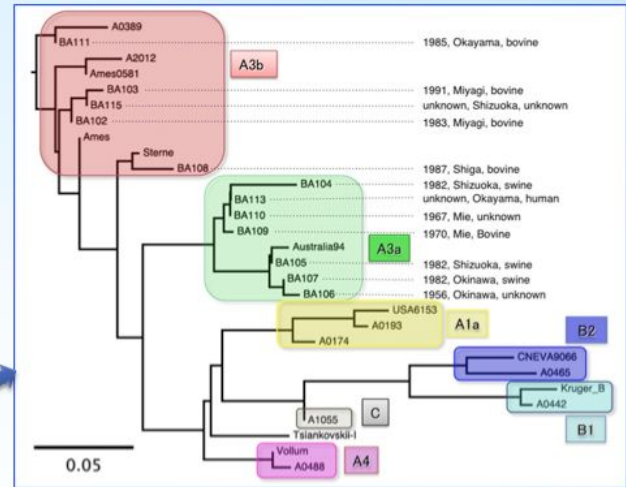


図6 今後、ペスト、野兎病、コクシエラ、類鼻疽、ボツリヌスと順々に構築予定である。