

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

1) 網羅配列解読による野兎病菌鑑別と病原性評価

いわき市立総合磐城共立病院・浅野先生から供試頂いた感染症の疑いのある不明病検体(右腋窩リンパ節膿瘍)を用いて病原体検索を試みた。

➤ 網羅配列解読のライブラリー作製:

膿瘍検体から解読用 DNA ライブラリーを作成するために、NEXTERA DNA sample prep kit (EpiCentre) を用いた。

➤ 次世代シーケンサーによる網羅塩基配列解読:

次世代シーケンサー illumina GAIIX で DNA ライブラリーの網羅解読を行った。解読キットは、TruSeq SBS kit v5 (illumine)を使用した。1 サンプルにつき 3,828 万本の解読リード (125 mer) を取得した。解読リードに内在するヒトゲノム配列を bwasw マッピングソフトで削除した。残った解読リードを用いて相同性検索 (megablast) を行い、病原体を検索した。現在公開されているほぼ全ての生物種の配列を内包している米国 NCBI nt 配列データベース(2011年7月11日版)に対して相同性検索を行い、病原体候補を検索した。得られた結果を MEGAN (v.4.40.6)にて類似性の見られた生物種の一覧図を得た(MEGANの閾値: 1 ヒット以上、score 200.0 以上)。

2) 1塩基バリエーション SNV による野兎病菌の株系統解析:

高病原性・野兎病菌タイプ A に属する *Francisella tularensis* subsp. *tularensis* SCHU_S4 株のゲノム情報 (GenBank ID: NC_006570.fna) をレファレンス配列にして、maq v0.7.1 によるリード・マッピングを行った。phred score, ≥ 20 ; coverage, ≥ 1 ; variant frequencies, ≥ 0.90 の条件で SNVs を抽出した (MapView 使用)。公開され利用可能な野兎病菌・10 株のゲノム情報を加え、該当 SNV サイトの

アレルを用いて最尤法による系統樹解析を行った (RAxML v7.25 を使用)。

3) バイオテロ対策のための次世代型網羅的病原体検索パイプラインの開発

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, 5, 6 を参照。

(倫理面への配慮)

微生物検査として網羅配列解読による病原体検出を行った。患者配列を特定する作業は行わず、個人情報に結びつく配列解析は一切行っていない。国立感染症研究所・医学研究倫理委員会にて承認済み。

C. 研究結果

1) 網羅配列解読による病原体鑑別と病原性評価

野兎病菌 (*Francisella tularensis*) は WHO 指定バイオテロ病原体 (カテゴリー A) に分類され、バイオテロ対策として臨床検体から迅速に野兎病菌の遺伝情報を得ることは重要である。本研究において、野兎病菌の感染症疑いの患者の腋窩部膿瘍 (図 1) を用いて、次世代シーケンサー illumina GAIIX による病原体の網羅的検出を行った。抗菌薬治療により膿瘍から起因菌の増殖を検出できなかったが、PCR 検査では野兎病菌陽性であった。

GAIIX にて 3,828 万本のリード (125 mer) を解読し、ヒト由来配列 (~99%) を完全に削除し、細菌と類似性を示す 833 本 (0.002%) の解読リードを得た (図 2)。相同性検索で得られた結果を taxonomy で分類したところ、全て野兎病菌 *Francisella tularensis* に相当する配列であった (図 3)。高病原性 Type A. *F. tularensis* subsp. *tularensis* と低病原性 Type B. subsp. *holarctica* との両方が約半数ずつヒットし、単純な相同性検索のみでは感染病原体の病原性を評価できなかった (図 3)。

野兎病菌 Type A (高病原性) と Type B (低病原

性)の双方を検出したことから、SNVs を利用した亜種を分類する株系統樹解析を行った。解読で得られた野兔病菌 833 リード(図3参照)を用いて、野兔病菌ゲノムに特徴的な SNV サイト 66 箇所を検出し、株固有の SNV アレルのアライメント解析を行った(図4)。Iwaki-08 は、日本分離株 FSC022 (Ebina) の SNV アレルと酷似しており、1 箇所のみ異なっていた (nt posi.: 329,375)。

アライメント結果を元に最尤法で系統樹を作成した。Iwaki-08 は日本固有に分布する *F. tularensis* subsp. *holarctica* (biovar *japonica*) と近縁であることが判明した。本症例は地域的な散发症例であり、バイオテロ使用が懸念される Type A (高病原性)とは明確に異なると鑑別することができた(図5)。

2) ベンチトップ型次世代シーケンサーの特徴

従来の次世代シーケンサーは大型シーケンサーという特徴から、大量配列の排出には有効である。しかしながら、バイオテロ発生等、有事に迅速に対応するためには解読時間の短縮が求められていた。大型シーケンサーよりも安価でかつ解読時間を短縮したベンチトップ型・次世代シーケンサーの登場により、従来よりも迅速に臨床検体の病原体検索が可能となった。昨年度に報告した野兔病菌感染症例を挙げると、大型シーケンサーでは最短7日間の解読・解析時間を要していたが、ベンチトップ型により2.5日にまで短縮することが可能になった。大型と比較して解読量は減るものの、臨床検体から病原体推定する場合には十分な量を得られることが分かった。

3) 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、感染研に検体を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6)。

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が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。

4) 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7)。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>

5) 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に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(図8)。

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。

6) コアゲノム SNPs を利用した菌種・菌株の類縁関係の特定

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9)。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こさぬよう、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。

D/E. 考 察・結 論

患者リンパ節膿瘍検体から網羅解読法で病原体・野兔病菌を検出し、その特徴的な株固有 SNVs を利用することで、病原性の強弱に関する遺伝情報を得ることができた。野兔病菌は病原性が異なる subsp. に分類され、北米由来の高病原性 Type A であるのか、低病原性 Type B による散発的な地域発生であるのか明確に分類する必要がある。これまでの分担研究におけるデータベース構築のもと、野兔病菌の網羅的なゲノム情報の収集とデータベース化を完了しており、本事例の解析を短期間で完了することができた。

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。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、誰もが簡単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

- Detection of a Possible Bioterrorism Agent, *Francisella* sp., in a clinical specimen using Next-generation Direct DNA Sequencing. Makoto Kuroda, Tsuyoshi Sekizuka, Fumiaki Shinya, Fumihiko Takeuchi, Takayuki Kanno, Tetsutaro Sata, Shigeyuki Asano. J Clin Microbiol. 2012, 50 (5): 1810-1812.
- 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

黒田誠、関塚剛史、竹内史比古。Deep sequencing 法による野兔病菌 *Francisella* 感染症の腋窩部膿瘍の網羅的病原体検索第 85 回日本細菌学会・総会 長崎市 2012 年 3 月

H. 知的財産権の出願・登録状況(予定を含む)

該当なし



図1 患者 (Iwaki-08) の臨床所見。A) 右親指、B) 右腋窩リンパ節、C) Bの拡大図

DNA extracted from clinical specimen (abscess)

↓ illumina DNA library by NEXTERA DNA library kit
125 mer direct DNA sequencing by GAIIX

38,285,502 reads (125 mer)

↓ Subtraction of human DNA sequence by bwasw mapping
(human genomic DNA sequences: hs_ref_GRch37.p2.fa)

48,223 reads (125 mer)

↓ blastn homology search against
non-redundant NCBI nt database
Taxonomic classification with MEGAN v4.40.6

Detection of *Fraxisella* sp. sequences (833 reads: 0.002%)

図2 Iwaki-08 患者の右腋窩リンパ節膿瘍からの病原体検索パイプライン

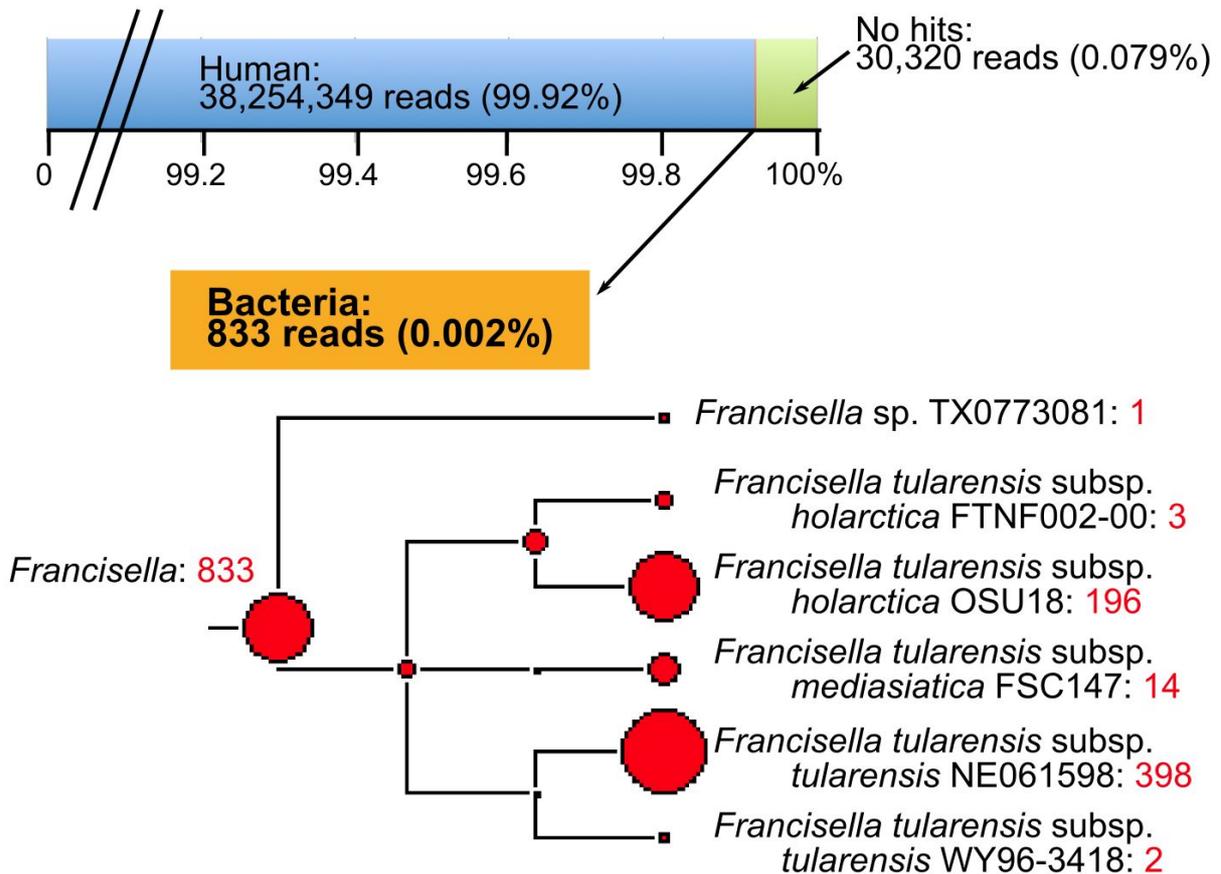


図3 イルミナ網羅解読にて取得した解読リードの解析。99.92%が膿瘍由来のヒト配列であり、細菌 bacteria と思われる配列は 833 リード (0.002%) ならずであった。ウイルス様配列は検出されなかった。細菌配列の相同性検索で得られた結果を taxonomy で分類したところ、全て野兔病菌 *Francisella tularensis* に相当する配列であった。*Francisella* の系統関係と subsp. そして、該当する系統のヒット・リード数 (赤字) で示している。高病原性 Type A. *F. tularensis* subsp. *tularensis* と低病原性 Type B. subsp. *holarctica* との両方が約半数ずつヒットし、単純な相同性検索のみでは感染病原体の病原性を評価できなかった。

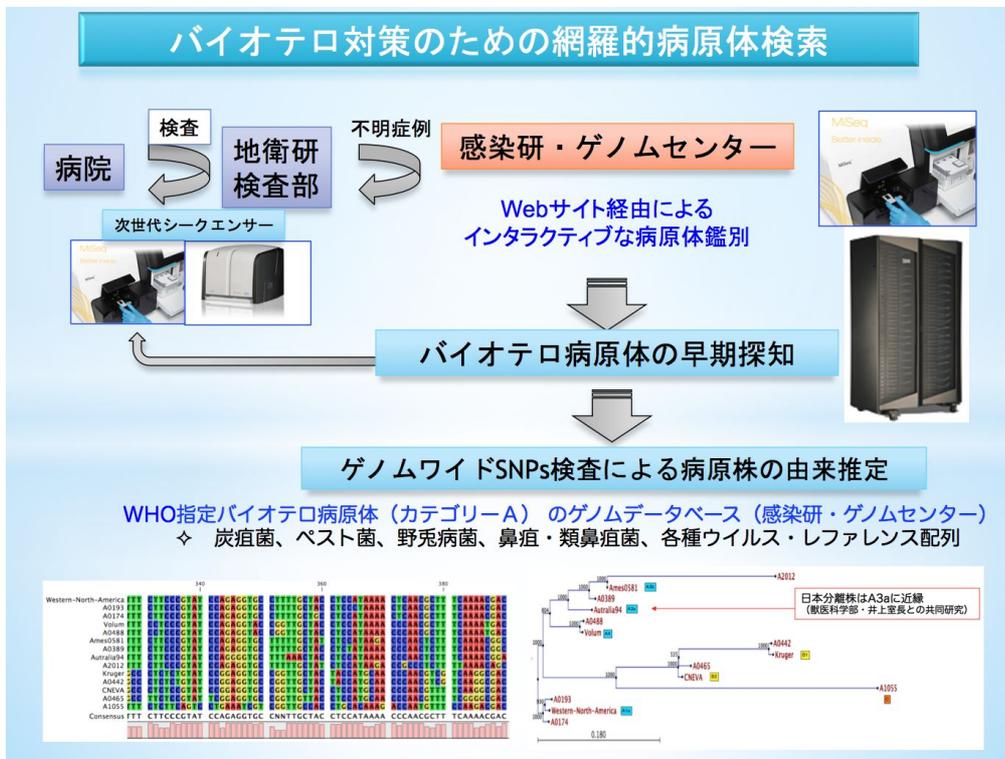


図6 ネットワーク経由によるバイオテロ病原体検索のための解析システム。現在、感染研・ゲノムセンターにベンチトップ型・次世代シーケンサーと情報解析サーバーが整備されている。地方衛生研究所等においても少しずつ整備されるようになってきた。検査技師等がインタラクティブに病原体鑑別を利用するために、Web interface による情報解析プログラムの開発と提供を試みた。

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>



Welcome to MePIC v2.0

Please login to start

User ID: Password:

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

[to MePIC manual](#)

- 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)

現在、アカデミアのみアカウント取得可能

図7 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。

MePIC v2.0

Log out
MePIC Manual

便 血液、髄液 咽頭ぬぐい液

MiSeq GS junior

ID: show

Project name

Select file(s) to upload
(Only .fastq.gz or .sff files, up to 2GBytes in total are acceptable.)

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。

Quality trimming and adaptor removal
Trimming params

5' trim length

Trim lower than this q-value

Quality threshold for fastq-mcf.pl

Minimum remaining sequence length

Host genome subtract
Subtract host genome

Database for subtraction

Homology search / read mapping

Search method

Database

E-value

Gap

Filter

Number of hits

Default

質の悪いリード、塩基を除去

ヒト配列を削除
(サル、マウスの選択も可能)

megablast検索 による核酸配列の照合

検索データ (.txt) をダウンロード
→ MEGAN v5 (チュービンゲン大学) のソフトで閲覧する

図8 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。

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:

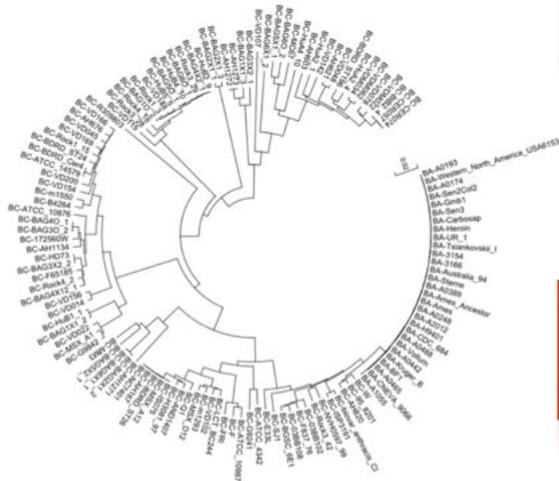
Input your sequence reads (Only .fastq.gz up to 2GBytes in total are acceptable.)

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。

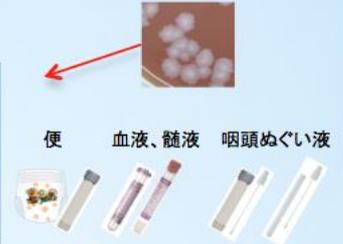
Send an e-mail after finished analyzing.
 Mail address:
 Change number of samples

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 \(compressed 13MBytes\)](#)



~~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9 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ー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。